Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 40

John von Neumann Institute for Computing (NIC)

From Computational Biophysics to Systems Biology (CBSB08)

edited by
Ulrich H. E. Hansmann
Jan H. Meinke
Sandipan Mohanty
Walter Nadler
Olav Zimmermann

Symposium, 19. - 21. May 2008
Forschungszentrum Jülich
Proceedings

organized by
John von Neumann Institute for Computing

NIC Series Volume 40

ISBN: 978-3-9810843-6-8

Die Deutsche Bibliothek – CIP-Cataloguing-in-Publication-Data
A catalogue record for this publication is available from Die Deutsche Bibliothek.

Publisher:	NIC-Directors
Technical Editor:	Martina Kamps
Distributor:	NIC-Secretariat Research Centre Jülich 52425 Jülich Germany
Internet:	www.fz-juelich.de/nic/
Printer:	Graphische Betriebe, Forschungszentrum Jülich

NIC Series Volume 40
ISBN: 978-3-9810843-6-8
Published in full on the internet
Persistent Identifier: urn:nbn:de:0001-00457
http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0001-00457

Preface

As in 2006 and 2007, the ”Computational Biology and Biophysics” group of the John von Neumann Institute for Computing (NIC) organized in 2008 again a workshop “From Computational Biophysics to Systems Biology” (CBSB08). Researchers and students from all over the world met from May 19th to May 21st, 2008 at the Research Centre Jülich and discussed physics-based approaches to systems biology. The increasing importance of computing in Biology, Medicine and other life science can be seen already by the number of participants which increased from 70 (CBSB06) to 150 (CBSB08) in only three years. As in the previous year, the participants explored in scientific presentations and numerous informal discussions a wide range of topics ranging from single macromolecules to the working of entire cells. This proceeding volume collects selected presentations from the 3-day long workshop that may serve as starting point for further discussions. It is divided into articles by invited speakers and such originally presented as posters or in contributed talks, as the interdisciplinary nature of the articles often defies a simple classification according to subject areas.

Besides the editors, Helga Frank, Erika Wittig and Martina Kamps helped organize the workshop. We also wish to thank IBM (Germany) for generous support.

Jülich, June 2008

U. H. E. Hansmann, J. H. Meinke, S. Mohanty, W. Nadler, O. Zimmermann

Contents

Invited Talks

Strategies to Overcome the Induced Fit Effects in Molecular Docking
I. Kufareva, R. Abagyan **1**

Simulating the Early Steps of Amyloid Fibril Formation and Disassembly
Ph. Derreumaux **7**

Milestoning
R. Elber, A. West **13**

Simulating Biomolecules in Cellular Environments
M. Feig, S. Tanizaki, J. Chocholoušová, M. Sayadi, J. W. Clifford, B. D. Connelly, S. Mukherjee, S. M. Law **23**

Scalable Systems for Computational Biology
Ch. Pospiech **31**

High Performance Computing in Multiscale Modeling Cardiac Contraction: Bridging Proteins to Cells to Whole Heart
J. J. Rice **37**

Determination of Protein Structure in Solution Based on $^{13}C^{\alpha}$ Chemical Shifts and NOE Distance Constraints
J. A. Vila, H. A. Scheraga **43**

Computer Simulation of Biomolecular Systems: Where Do We Stand?
W. F. van Gunsteren, D. P. Geerke **49**

Towards Accurate Predictions of Binding Affinities: The Simulated Scaling Based Free Energy Method
L. Zheng, H. Li, W. Yang **57**

Contributed Talks

Protein-Ligand Docking Including Protein Flexibility: A Hierarchical Approach
I. Antes, Ch. Hartmann, Th. Lengauer **65**

Electroporation Studied by Molecular Dynamics Simulations
R. A. Böckmann, S. W. I. Siu, S. Leis, B. L. de Groot, S. Kakorin, E. Neumann, H. Grubmüller **69**

Modeling Conformational and Molecular Weight Heterogeneity with Analytical Ultracentrifugation Experiments
B. Demeler, E. Brookes **73**

Exploring the Free Energy Surface of Short Peptides by Using Metadynamics
C. Camilloni, A. De Simone **77**

Association of Neamine and its Derivative with the Ribosomal A-Site RNA
M. Długosz, J. Trylska **81**

Simplified Approaches to Complex Biological Systems
N. V. Dokholyan **85**

Predicting Protein Interactions from Functional Specificity
K. A. Feenstra, G. Bastianelli, J. Heringa **89**

Short Membrane Proteins from Viruses: Channel-Pore Dualism?
G. Patargias, H. Martay, J. Krüger, C.-P. Chen, Y.-T. Wang, M.-D. Lee, C.-C. Chen, G.-S. Hong, W. B. Fischer **93**

Coarse-Grained Molecular Models for High-Throughput and Multi-Scale Functional Investigations
K. Hamacher **97**

Simulation of RNA Folding on the Simple Cubic Lattice
S. Hayryan, K. G. Sargsyan,C.-K. Hu **101**

Designing an Automatic Pipeline for Protein Structure Prediction
S. Kmiecik, M. Jamroz, A. Zwolinska, P. Gniewek, A. Kolinski **105**

Conformational Polymorphism of a Fibrillogenic Fusion Peptide in Explicit Solvent and at an Interface
V. Knecht, R. Lipowsky **109**

Modeling the Free Energy of Polypeptides in Different Environments
G. La Penna, S. Furlan, A. Perico **113**

Protein-Ligand Docking with Two-Scale Receptor Dynamics and QM/MM Potential
J. A. Kalinowski, B. Lesyng **117**

Inhibition of Fibril Formation of Beta-Amyloid Peptides
N. S. Lam, M. Kouza, H. Zung, M. S. Li **121**

A Topological Model of a "Jumping Gene" Machine
J. Navarra-Madsen **125**

Computer Modeling of Small Ligands Diffusion in Drosophila Melanogaster Hemoglobin
L. Dams, S. Orlowski, W. Nowak **129**

High Throughput In-Silico Screening Against Flexible Protein Receptors
H. Sánchez, B. Fischer, H. Merlitz, W. Wenzel **133**

Mutations as Trapdoors: The Rop-Dimer with Two Competing Native Conformations
A. Schug, J. Onuchic **137**

Global Motions in the Nucleosome Explored Using a Coarse-Grained Model
K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. C. Smith, J. Langowski **141**

Effects of Confinement and Crowding in Protein Folding
M. Wojciechowski, M. Cieplak **145**

Posters

Conformational Study of Alzheimer's Aβ Wild Type Peptide and Flemish Mutant
P. Anand, F. S. Nandel, U. H. E. Hansmann **149**

Effect of Temperature on the Structural and Hydrational Properties of Human Islet Amyloid Polypeptide in Water
M. N. Andrews, I. Brovchenko, R. Winter **153**

A Knowledge-Based Potential for Protein-RNA Docking
R. P. Bahadur, M. Zacharias **157**

Fast and Accurate Structure-Based Calculation of Folding Free Energies and Binding Affinities
C. M. Becker, A. Benedix, B. L. de Groot, A. Caflisch, R. A. Böckmann **161**

Correlation Effects in Protein-Protein Recognition
H. Behringer, F. Schmid **165**

Prediction of Twist of Amyloid Fibrils Using Molecular Dynamics
J. Berryman, S. Radford, S. Harris **169**

Automatic Sequential NOESY Assignment and NMR Structure Improvement by X-Ray
K. Brunner, W. Gronwald, A. Fischer, J. Trenner, K.-P. Neidig, H. R. Kalbitzer **173**

Exploring the First Steps of Aβ16-22 Protofibril Disassembly by N-Methylated Inhibitors
Y. Chebaro, Ph. Derreumaux **177**

Simulation of the Outer Membrane Protein X in a Lipid Bilayer and in a Micelle
A. Choutko, A. Glättli, W. van Gunsteren **181**

OPERA: An OPtimized coarse-grained Energy model for RnA
C. Colas, P. H. Nguyen, J.-Ch. Gelly, Ph. Derreumaux **185**

Understanding of High Pathogenicity of the Avian Influenza Virus H5N1: Why H5 is Better Cleaved by Furin
P. Decha, P. Wolschann, S. Hannongbua **189**

Bound Water as a Tool to Detect Soluble Amyloid Oligomers and Amyloid Protofibrils, the Early Stage of Development of the Alzheimer's Disease
F. Despa **193**

Stabilizing Regions in Membrane Proteins
F. Dressel, A. Marsico, A.Tuukkanen, R. Winnenburg, M. Schroeder, D. Labudde **197**

Loop Parameterization and RNA Secondary Structure Folding
T. R. Einert, P. Näger, H. Orland, R. R. Netz **201**

Investigation of Structure of Amphipathic Peptides in Different Environments via Replica Exchange Molecular Dynamics Simulations
Ö. Engin, M. Sayar **205**

Protein-Protein Interaction Prediction
F. Fink, St. Ederer, W. Gronwald **209**

Determining RNA Flexibility by Graph Theory: Ribosomal Exit Tunnel as a Case Study
S. Fulle, H. Gohlke **213**

***Ab initio* Molecular Dynamics of the Zn–Binding Site of the Alzheimer's Amyloid β–Peptide**
S. Furlan, G. La Penna **217**

Do Amyloidogenic Regions Intersect with Folding Nuclei of Native Structure?
O. V. Galzitskaya **219**

Coarse-Grained Simulations of Protein Adsorption on Solid Surfaces
B. Griepernau, Ch. Hanke, L. Santen, R. A. Böckmann **223**

Towards Understanding the Early Events in the Conformational Transition of Amyloid Beta Peptides
S. Hayat, V. Helms **227**

Protein Interactions with their Environment
M. Hoefling, K. E. Gottschalk **231**

Identification of Differential Protein Expression in Response to the Application of BioRegulators that Enhance Plant Productivity and Quality
R. Horn, D. Zimmermann, St. Schillberg **235**

Hamiltonian Replica Exchange Molecular Dynamics Using Soft-Core Interactions to Enhance Conformational Sampling
J. Hritz, Ch. Oostenbrink **237**

TollML: A Database of Toll-Like Receptor Structural Motifs
F. Jamitzky, J. Gong, T. Wei, W. M. Heckl, S. C. Rössle **241**

The Glucan, Water Dikinase - A Kinetic Model to Understand the Initial Step in Starch Mobilization in Plant Leaves
Ö. Kartal, O. Ebenhöh **245**

Engrailed Homeodomain Folds Overnight by 100 Processors
K. V. Klenin, W. Wenzel **249**

Single Molecule FRET Study of the Conformational Energy Landscape of a Transfer RNA
A. Yu. Kobitski, M. Hengesbach, M. Helm, G. U. Nienhaus **253**

Impact of Induced Fit on Ligand Scoring and a Strategy of Identifying a Minimal Set of Flexible Residues
D. Kokh, W. Wenzel **257**

Contributions to the Hydration Free Energies of Amino Acids
G. König, St. Boresch **261**

Conservation Analysis of Functional Important Residues of the Oxygen Evolving Mechanism Located in the D1 Subunit of Photosystem II
E.-M. Krammer, G. M. Ullmann **265**

Viral Membrane Proteins: Flexibility and Assembly
J. Krüger, W. B. Fischer **269**

Hen Egg White Lysozyme Adsorption on a Mica Surface: A Fully Atomistic Molecular Dynamics Study
K. Kubiak, P. Mulheran **273**

Internal Dynamics of Ribosomal Elongation Factors G and Tu Studied with All-Atom and Coarse-Grained Molecular Dynamics
K. Kulczycka, M. Długosz, J. Trylska **277**

Efficient Molecular Docking of Drug Molecules into DNA Targets and their Enrichment by Cutting-Edge Technologies
S. A. Kumar, M. Petersen, M. G. B. Drew, M. M.-L. Grage, C. J. Cardin, Y. Gan **281**

Applications of a Novel Biasing Potential to Study DNA Translocation and DNA Base Flipping
S. M. Law, M. Feig **285**

Characterization of the Quinolone-Gyrase-Interaction Using Docking, Molecular-Dynamics and Site-Directed Mutagenesis
J. Lenz, Th. Lemcke, P. Heisig, A. Torda **289**

Classification of Kinases: A Fast, Automated Structure-Based Approach
J. Lenz, Th. Margraf, Th. Lemcke, A. Torda **293**

Modelling of Possible Binding Modes of Caffeic Acid Derivatives to JAK3 Kinase
J. Kuska, P. Setny, B. Lesyng **297**

Looking for Inhibitors of RIO Kinases
M. Geller, Ł. Walewski, M. Długosz, B. Lesyng **301**

1.4-DHP-Lipid Forms a Tubular Micellae
I. Liepina, C. Czaplewski, V. Ose, R. Danne, G. Duburs **305**

Protein Structure Prediction Using Coarse Grain Force Fields
N. Mahmood, A. Torda **309**

HANSWURST: Fast Efficient Multiple Protein Structure Alignments
Th. Margraf, A. Torda **313**

Computer Simulations of Product Dissociation from the Active Site of the Anthrax Edema Factor
L. Martínez, Th. Malliavin, M. Nilges, A. Blondel **317**

A Software Library for Monte Carlo-Based Rigid Body Modelling Against Small Angle Scattering Data
Ch. Meesters **321**

Development of a Docking Methodology for Predicting the Structure of Protein-Protein Complexes
I. Meliciani, S. M. Gopal, W. Wenzel **325**

A Non-Native Helix Extension Channels Folding
S. Mohanty, U. H. E. Hansmann **329**

Modulation of Aggregate Size and Shape Distributions of Amyloid-β Peptide Solutions by a Designed β-Sheet Breaker
L. Nagel-Steger, B. Demeler, K. Hochdörffer, Th. Schrader, D. Willbold **333**

Flexible Peptide-Protein Docking Employing PSO@Autodock
V. Namasivayam, R. Günther **337**

Exploration of the Energy Landscape of Protein-Protein and Antibody-Antigen Interactions
J. Neumann, J. Morfill, K. E. Gottschalk **341**

From Isotope Labeling Patterns to Metabolic Flux Rates
K. Nöh, M. Weitzel, W. Wiechert **345**

The Inherent Stability of Collagen
V. K. Pálfi, A. Perczel **349**

Molecular Dynamics Simulations of the Metaloenzyme Thiocyanate Hydrolase with Non-Corrinoid Co(III) in Active Site
L. Peplowski, W. Nowak **353**

Constraint Network Analysis: Exploiting the Link Between Protein Rigidity and Thermostability
S. Radestock, H. Gohlke **357**

Solvent in Protein Interfaces: Molecular Dynamics Approach
S. Samsonov, J. Teyra, M. T. Pisabarro **361**

Nearly-Deterministic Methods for Optimising Protein Geometry
G. Schenk, A. Torda **365**

Different Types of Protein Folding Identified with a Coarse-Grained Heteropolymer Model
St. Schnabel, M. Bachmann, W. Janke **369**

Effect of Surfaces on the Aggregation of Hydrophobic and Hydrophilic Amyloidogenic Peptides
G. Singh, I. Brovchenko, R. Winter **373**

Free Energy Study of Ion Permeation through Gramicidin
S. W. I. Siu, R. A. Böckmann **377**

Free-Energy Based All-Atom Protein Folding Using Worldwide Distributed Computational Resources
T. Strunk, S. M. Gopal, K. Klenin, W. Wenzel **381**

Parallel 2d-Wavelet Transform on the Cell/B.E. for Fast Calculation of Coulomb Potentials
A. Schiller, G. Sutmann **385**

RedMD – A New Package for Reduced Molecular Dynamics
M. Szypowski, A. Górecki, J. Trylska **389**

Insights into the Self-Assembly of Phenylalanine Oligopeptides by Replica Exchange MD Simulations with the GBSW Implicit-Solvent Model
P. Tamamis, L. Adler-Abramovich, E. Gazit, G. Archontis **393**

Analysis and Classification of the Structural Interactome
J. Teyra, M. Paszkowski-Rogacz, G. Anders, M. T. Pisabarro **397**

Characterization of the Binding Surface of the Human Protein GABARAP
Y. Thielmann, O. H. Weiergräber, J. Mohrlüder, B. W. Koenig, R. Hartmann, Th. Stangler, K. Wiesehan, D. Willbold **401**

Freezing and Collapse of Flexible Polymers
Th. Vogel, M. Bachmann, W. Janke **405**

A Computational Approach to Study the Energy Transduction Mechanism in the Na^+/K^+-ATPase
Ch. Weidemüller, K. Hauser **409**

Improved Hydrogen Bond Potentials to Assess *in Silico* Protein Folding
M. G. Wolf, J. Vreede, P. G. Bolhuis, S. W. de Leeuw **413**

Sequence Specific DNA Recognition by EcoRV
M. Zahran, P. Imhof, J. C. Smith **417**

Workshop Program **421**

List of Participants **425**

Strategies to Overcome the Induced Fit Effects in Molecular Docking

Irina Kufareva and Ruben Abagyan

The Scripps Research Institute, La Jolla, CA 93037, USA
E-mail: {*kufareva, abagyan*}*@scripps.edu*

Protein flexibility and induced fit effects present a major obstacle to the development of better molecular docking algorithms and scoring functions. Here we present several modeling and sampling strategies to generate high quality ligand-induced receptor conformations. We demonstrate that the repulsive-density approach, the alanine-scanning (SCARE) algorithm, and the ligand-guided receptor modeling significantly improve the ligand pose prediction, de novo inhibitor finding via virtual screening, and ligand profiling results.

1 Introduction

Recent years have been characterized by great advances of structural proteomics and exponential growth of the number of newly solved protein structures. However, these structures still represent only a small fraction of therapeutically relevant proteome, with other proteins only accessible in the form of approximate homology models. Moreover, as static snapshots of protein dynamic flexibility, they are often not fully compatible with ligands of interest. Ligand docking, screening and profiling efforts invariably fail in such cases.

General structure prediction methods of molecular dynamics simulations partially address the problem, however, they are still incapable of identifying a reasonably small set of dockable and screenable ligand binding pocket conformations. Alternatively, concurrent pocket/ligand global optimization can provide an exhaustive solution to the problem, however, it is impractical due to time and resource requirements. A reasonable alternative is the so-called mutiple receptor conformation (MRC) docking[1]. In this approach, the protein flexibility is represented by a series of rigid snapshots. For the best experimentally explored cases, experimental snaphots (e.g. multiple crystallographic structures or NMR ensembles) can be used as input for the MRC docking.

In the absence of mutiple experimental structures, *ab initio* or ligand-guided simulations are needed to computationally generate an ensemble of receptor conformations. Available approaches to this task include molecular dynamics (e.g. Eyrisch and Helms[2]), normal mode analysis (e.g. Cavasotto, Kovacs and Abagyan[3]), and internal coordinate sampling (e.g. Abagyan and Totrov[4]). All these methods, however, tend to produce far too many models, most of them inappropriate for ligand docking due to insuficient volume and shape of the binding pocket. Furthermore, the large number of pocket conformers tends to decrease, rather than increase, the number of false positives in both pose prediction and compound scoring and screening. In addition the large conformational ensembles quickly become overwhelming for docking algorithms. Therefore, it is important to either compress the generated conformational ensemble or reliably select its most informative representatives.

We here present several modeling and sampling strategies to generate high quality ligand-induced receptor conformations. The first of them allows generation of more drug-

gable pocket models by introducing a pocket controlling device in the form of repulsive density. This method and its successful application to protein kinase CK2 are described in Section 2. In some cases, while a major fraction of a crystallograhic or generated pocket very closely resembles its bound conformation, the structural rearangements in the remaining fraction creates steric hindrances with the ligand and makes it impossible to reproduce the correct ligand binding geometry. In such cases, computational removal of the incorrectly placed elements provides a partially correct/partially empty pocket with sufficient number of native contacts for the ligand to dock in the near-native pose. The common wisdom underlying this approach is "better no atom than a wrong atom". This aproach was successfully implemented in the *SCan Alanines and REfine* (SCARE) flexible receptor docking algorithm presented in Section 3. Computational excision, rather than modeling, of structural elements of the receptor is also beneficial in cases where ligands bind to pockets previously occupied by these elements. The so-called type II inhibitors of protein kinases form a well-studied family of such interactions. In Section 3, we also present DOLPHIN (*Deletion Of Loop from PHe-IN*) kinase models as powerful devices for ligand development, screening, and activity profiling.

Known strong small molecule binders to the protein of interest can provide a valuable information for both generation and selection of receptor pocket conformations in MRC docking, and help reduce the complexity and dimensionality of the problem. One way to employ this information is a so-called ligand-guided receptor selection, where among the mupltiple conformations, only those are chosen for screening that are selective towards known ligands. An overview of several successful applications of this strategy is described in Section 4.

Overall, the described approaches to the MRC generation and selection, along with proper methods of ligand docking and scoring, represent a great advance in the field of induced fit docking.

2 *Ab Initio* Pocket Ensemble Generation

A common drawback of simulation methods is the large number of generated models, with only a small fraction of them being compatible with ligand binding. Clearly, there is a need in computational techniques that can successfully guide the simulation procedures towards more druggable pockets. Experimental information in the form of known ligands targeting the pocket and their activities may be of great help. However, for *de novo* pockets such information is not available. We developed a computational protocol that addresses this problem by introducing a repulsive density in the pocket as an independent energy/penalty term. Such density represents a generic ligand and prevents the simulation procedure from generating conformations in which elements of the structure collapse inside the pocket. The algorithm of density generation includes the following steps:

1. Identification of residues whose sidechains form the pocket.
2. Simultaneous conversion of these residues to Ala.
3. Construction of atom density grid map for the obtained protein.
4. Repeated spatial averaging of the map in order to obtain a smoothed density cloud filling the cavities of the original protein.

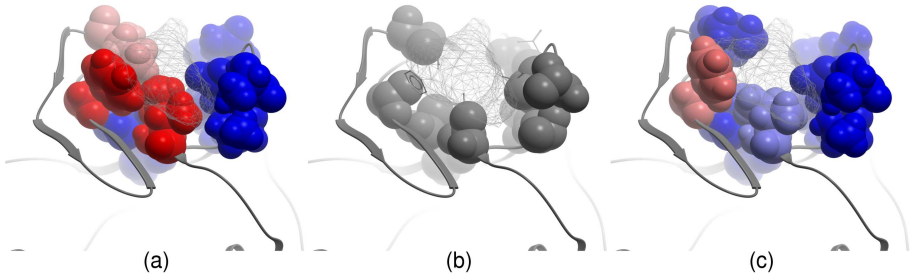


Figure 1. Repulsive density blob represents a generic ligand and guides side-chain sampling procedure towards more open pocket conformations. (a) - Initial closed pocket conformation, (b) - pocket sidechain alanine mutation and density generation, (c) - resulting open pocket conformation.

5. Taking a difference of the smoothed and original maps.

The procedure results in blobs of density filling in the cavities of the protein. Due to side-chain alanine conversion, the generated density represents the maximal volume of the pocket achievable without backbone rearrangements. The density is further used as an additional energy/penalty term in montecarlo side-chain simulation (Fig. 1).

This protocol was applied for *de novo* finding of compounds that bind to the N-terminal lobe of protein kinase CK2 and prevent its interaction with the regulatory subunit CK2β. Screening a large virtual chemical database against the generated ensemble of four most druggable conformations yielded a series of compounds that were experimentally validated and confirmed to inhibit the subunit interaction in a dose-dependent manner.

3 Better Deleted Than Displaced

The generated receptor conformations will rarely cover a 100% of the protein pocket conformational space. It is safe to assume that any generated conformation is only partially correct, with some elements still interfering with ligand binding. Therefore, a customized or systematic removal of those parts may lead to “dockable” pocket models in which the ligand binding geometry is successfully reproduced, provided that the removed parts is not the main determinant of the ligand-receptor interaction. This part can later be brought in as a part of the refinement, if needed.

3.1 SCARE

The majority of the induced fit changes involve a few protein side-chains and only a minor adjustment of the backbone. Therefore removal the pocket parts at the side-chain level is frequently sufficient. We developed an algorithm that systematically scans pairs of neighboring side chains in the binding pocket, replaces them by alanines, and docks the ligand to each “gapped” version of the pocket. All docked positions are scored, refined with original side chains and flexible backbone and re-scored. The optimal SCARE (SCan Alanines and

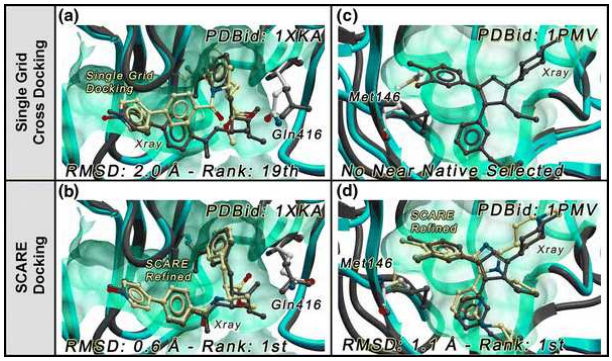


Figure 2. The SCARE algorithm successfully reproduces ligand binding geometry in difficult cases where traditional rigid-receptor docking fails.

REfine)[5] protocol identifies a near native conformation (under 2 A RMSD) as the lowest rank for as many as 90% of the cross-docking complexes (compare to 50% success rate with optimal single receptor cross-docking) (Fig. 2). The procedure predicts not only the binding pose of a ligand, but also conformational changes induced by its binding, therefore producing a new highly relevant protein conformation that can be used in VLS along with the original one.

3.2 DOLPHIN

In the recent years, there is a great interest to a specific type of protein kinase inhibitors, the so-called *type II inhibitors* that induce a transition of the kinase activation loop from its active, DFG-in, position, to the DFG-out state. This transition is too large to be modeled by any existing computational method. However, we found that the above "better deleted than displaced" strategy is very helpful in induced fit docking of type II ligands to DFG-in structures. We showed that *Deletion-Of-Loop from PHe-IN* (DOLPHIN, Fig. 3) kinase modification leads to models which

1. Reproduce the correct binding geometry of the existing type II ligands.
2. Selectively score active type II ligands higher than inactives and decoys.
3. Provide means for *in silico* ligand activity profiling.

Fig. 4 illustrates a great potential of the DOLPHIN kinase models as screening and profiling devices for type II kinase inhibitors.

4 Ligand Guided Model Selection and Generation

Multiple receptor conformations for high throughput ligand docking can be generated with one or several ligands actually present in the binding site. Fully receptor-flexible docking of a few known ligands can be performed to force the receptor into alternative conformations. The irrelevant generated conformations can then be filtered out by evaluating

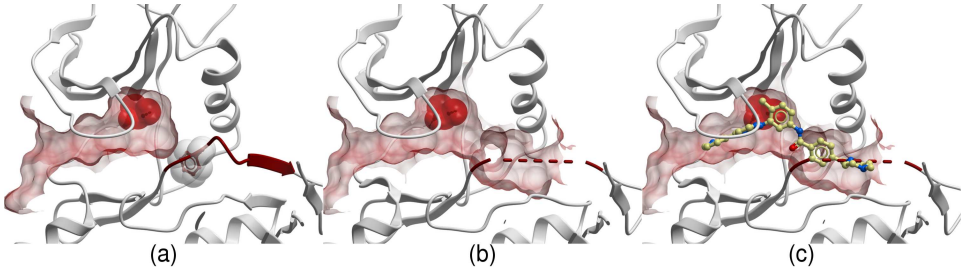


Figure 3. The DOLPHIN kinase models are powerful screening devices for type II kinase inhibitors. (a) - Initial DFG-in structure, (b) - modified structure, (c) - docking of a type-II ligand.

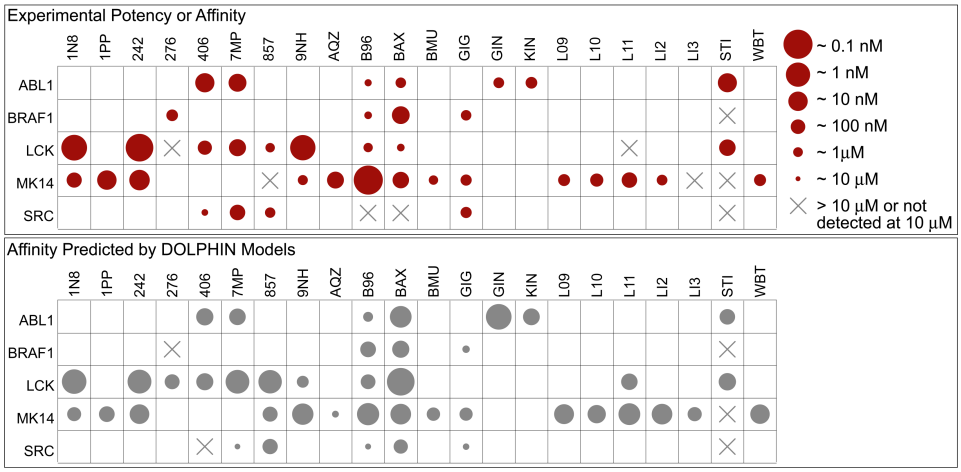


Figure 4. Comparison of experimental activities of known type II kinase inhibitors with their binding affinities predicted by the DOLPHIN kinase models.

the enrichment factor on a test set. Bisson et al.[6] generated multiple conformations of androgen receptor with two different antagonists by Monte-Carlo sampling in ICM. Each conformation was tested for its ability to discriminate between AR binders and non-binders in a panel of 88 nuclear receptor ligands. The two AR conformations with the best enrichment characteristics were then used for virtual ligand screening of the marketed drugs for potential androgen receptor antagonists. Three identified antipsychotic drugs exhibited anti-androgenic activity and were then rationally re-purposed to nonsteroidal molecules with improved AR antagonism and marked reduction in affinity for dopaminergic and serotonergic receptors. Similar procedure led to productive models of two G-protein coupled receptors, M2 muscarinic receptor and melanin concentrating hormone receptor[7], and resulted in successful *de novo* indentification of new modulator chemotypes.

5 Conclusion

We described a series of computational approaches that allow to model receptor flexibility in molecular docking. These methods present a practical alternative to concurrent protein and ligand simulation, and lead to productive receptor models suitable for ligand docking, screening, and profiling. The new approaches were successfully applied to find novel inhibitors in several difficult cases.

Acknowledgments

This work was funded by NIH/NIGMS grants 5-R01-GM071872-02 and 1-R01-GM074832-01A1.

References

1. M. Totrov and R. Abagyan, *Flexible ligand docking to multiple receptor conformations: a practical alternative*, Curr. Opin. Struct. Biol., **18**, 178–184, Apr 2008.
2. S. Eyrisch and V. Helms, *Transient pockets on protein surfaces involved in protein-protein interaction*, J. Med. Chem., **50**, 3457–3464, Jul 2007.
3. C.N. Cavasotto, J.A. Kovacs, and R.A. Abagyan, *Representing receptor flexibility in ligand docking through relevant normal modes*, J. Am. Chem. Soc., **127**, 9632–9640, Jul 2005.
4. R. Abagyan and M. Totrov, *Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins*, J Mol Biol, **235**, no. 3, 983–1002, Jan 1994.
5. G. Bottegoni, I. Kufareva, M. Totrov, and R. Abagyan, *A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE)*, J. Comput. Aided Mol. Des., **22**, 311–325, May 2008.
6. W.H. Bisson, A.V. Cheltsov, N. Bruey-Sedano, B. Lin, J. Chen, N. Goldberger, L.T. May, A. Christopoulos, J.T. Dalton, P.M. Sexton, X.K. Zhang, and R. Abagyan, *Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs*, Proc. Natl. Acad. Sci. U.S.A., **104**, 11927–11932, Jul 2007.
7. C.N. Cavasotto, A.J. Orry, N.J. Murgolo, M.F. Czarniecki, S.A. Kocsi, B.E. Hawes, K.A. O'Neill, H. Hine, M.S. Burton, J.H. Voigt, R.A. Abagyan, M.L. Bayne, and F.J. Monsma, *Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening*, J. Med. Chem., **51**, 581–588, Feb 2008.

Simulating the Early Steps of Amyloid Fibril Formation and Disassembly

Philippe Derreumaux

Laboratoire de Biochimie Théorique, UPR 9080 CNRS, IBPC, Université 7 Paris-Diderot,
13 rue Pierre et Marie Curie, 75005, Paris, France
E-mail: philippe.derreumaux@ibpc.fr

More than 20 human diseases are associated with the self-assembly of proteins into transient cytotoxic oligomers and eventually amyloid fibrils. Alzheimer's disease, affecting today more than 15 million people world-wide, is characterized by the aggregation of the Aβ(1-40)/Aβ(1-42) peptides. Because aggregation is very complex, structural characterization of the intermediate species remains to be determined. Similarly, though N-methylated Aβ(16-22) peptides inhibit the fibrillogenesis of full-length Aβ and disassemble fibrils $in\ vitro$, there is little information about their mechanism of action. Here, I review recent coarse-grained protein simulations aimed at understanding the dynamics and free energy surface of amyloid-forming peptides using the activation-relaxation technique, molecular dynamics and replica exchange molecular dynamics simulations.

1 Introduction

Neurodegenerative diseases such as Alzheimer's and prion diseases are characterized by the aggregation of non-related proteins of various amino acid lengths and compositions.[1] There is strong evidence that the soluble oligomers, forming in the early steps of aggregation, are the most cytotoxic species.[2] Structural characterization of these species is difficult, however, experimentally, because they are unstable and span a timescale of several days $in\ vitro$. One numerical challenge in characterizing these transient oligomers is the development of coarse-grained models and sampling methods able to explore, at an appropriate atomic resolution, large time and spatial scales. Recently, Ma and Nussinov gave a schematic overview over some simulation methods, including all-atom molecular dynamics (MD) simulations in explicit solvent and coarse-grained DMD simulations.[3] All-atom implicit solvent simulations[4–6] and coarse-grained Langevin dynamics[7] are also performed. Here I present recent simulations aimed at understanding the dynamics and free energy surface of amyloid-forming peptides using a coarse-grained implicit solvent protein force field (OPEP) coupled to the activation-relaxation technique (ART), MD and replica exchange MD (REMD) simulations. In an accompanying paper, we present coarse-grained MD and REMD simulations of Aβ(16-22) oligomers with multiple copies of an N-methylated inhibitor.

The coarse-grained protein simulation package is discussed in Sec. 2, with emphasis on the force field and the benchmarks used for validating the protein model. The results of protein aggregation simulations will be outlined in Sec. 3 and Sec. 4.

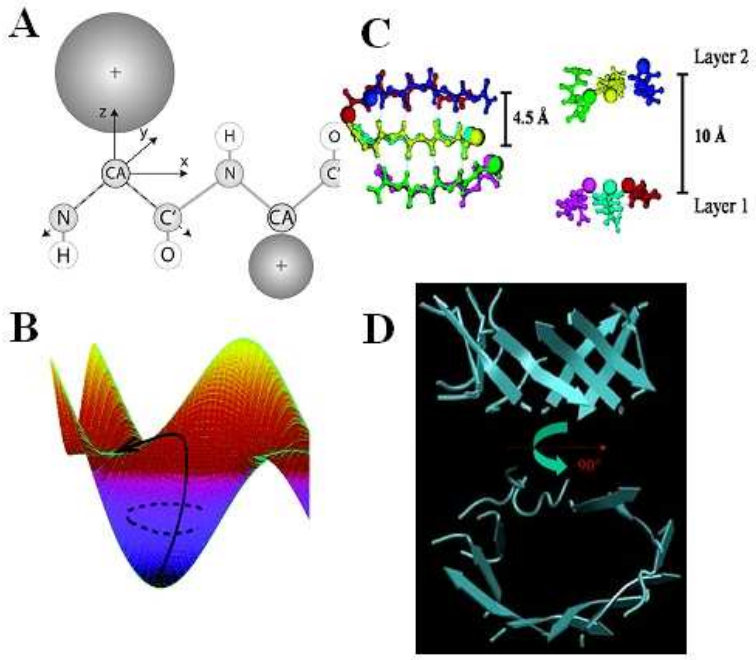


Figure 1. (A) Schematic picture of the coarse-grained model. Most side-chains are represented by a sphere with appropriate radius and physico-chemical property. (B) Sketch of the activation-relaxation technique. Starting from a minimum, the system is subject to a random perturbation and is pushed along the eigenvector of negative eigenvalue until a saddle point is reached, then relaxed to the connecting minimum. (C) A local energy minimum found by ART and MD simulations on Aβ(16-22) hexamer displaying structural characteristics of cross-β structure. (D) A representative β-barrel found in MD-OPEP simulations of 16 β2m(83-89) chains starting from a disordered state.

2 The Coarse-Grained Protein Simulation Package

2.1 OPEP Force Field

The off-lattice model we use consists in a detailed representation of the backbone, modelled by its N, H, Cα, C', O atoms and in one bead or centroid for all side-chains, except the proline amino acid which is represented by all heavy atoms (see Fig.1A). OPEP (Optimized Potential for Efficient structure Prediction) version 3.2 is expressed as a sum of local, nonbonded (VdW) and hydrogen-bond (H-bond) terms:[8]

$$E_{local} = \sum_{bonds} K_b(r - r_{eq})^2 + \sum_{angles} K_\alpha(\alpha - \alpha_{eq})^2 + \sum_{imp-tors} k_\Omega(\Omega - \Omega_{eq})^2 + \sum_{\phi} k_{\phi\psi}(\phi - \phi_o)^2 + \sum_{\psi} k_{\phi\psi}(\psi - \psi_o)^2 \quad (1)$$

The term E_{local} contains force constants associated with changes in bond lengths and bond angles of all particles, changes in improper torsions of the side-chains and the peptide

bonds and changes in ϕ and ψ angles, where $\phi_0 = \phi$ within the interval $[\phi_{\rm lower}, \phi_{\rm upper}]$ and $\phi_0 = \min(\phi - \phi_{\rm lower}, \phi - \phi_{upper})$, otherwise, with $\phi_{\rm lower} = -160°$, and $\phi_{\rm upper} = -60°$, respectively. Similarly, we use $\psi_{lower} = -60°$ and $\psi_{upper} = 160°$. Note this analytic form does not prevent sampling of conformations covering all values of ϕ and ψ.

The nonbonded interactions are modelled by:

$$E_{VdW} = \epsilon_{ij} \left(\left(\frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} \right) H(\epsilon_{ij}) - \epsilon_{ij} \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} H(-\epsilon_{ij}) \qquad (2)$$

Here the Heavyside function $H(x) = 1$ if $x \geq 0$ and 0 if $x < 0$, r_{ij} is the distance between the particles i and j, r^0_{ij} = (r^0_i + r^0_j)/2 with r^0_i the Van der Waals radius of particle i.

Finally, the E_{H-bond} term consists of two-body (E_{HB1}) and four-body (E_{HB2}) terms:

$$E_{HB1} = \sum_{ij, j=i+4} \varepsilon_{hb1-4}\, \mu(r_{ij}) \nu(\alpha_{ij}) + \sum_{ij, j>i+4} \varepsilon_{hb1>4}\, \mu(r_{ij}) \nu(\alpha_{ij}) \;\; with \qquad (3)$$

$$\mu(r_{ij}) = 5(\frac{\sigma}{r_{ij}})^{12} - 6(\frac{\sigma}{r_{ij}})^{10} \qquad (4)$$

$$\nu(\alpha_{ij}) = \begin{cases} \cos^2 \alpha_{ij} \;,\; \alpha_{ij} > 90° \\ 0, \; otherwise \end{cases} \qquad (5)$$

The sum in eq. (3) is over all residues i and j separated by $j = i + 4$ and $j > i + 4$ (helices 3_{10} are thus excluded), r_{ij} is the O..H distance between the carbonyl oxygen and amide hydrogen, α_{ij} the NHO angle and σ, set to 1.8Å, the equilibrium value of the O..H distance.

Four-body effects, which represent cooperative energies between hydrogen bonds ij and kl, are defined by

$$\begin{aligned} E_{HB2} = & \sum \varepsilon^{coop}_{\alpha} \exp(-(r_{ij} - \sigma)^2/2) \; \exp(-(r_{kl} - \sigma)^2/2) \; \Delta(ijkl) \\ & + \sum \varepsilon^{coop}_{\beta} \exp(-(r_{ij} - \sigma)^2/2) \; \exp(-(r_{kl} - \sigma)^2/2) \; \Delta'(ijkl) \end{aligned} \qquad (6)$$

with $\varepsilon^{coop}_{\alpha}$ and $\varepsilon^{coop}_{\beta}$ the cooperative energies for α-helices and β-sheets. The parameter $\Delta(ijkl)$ is set to 1 if residues (k, l) = (i+1, j+1) and ($j = i$+4, $l = k+4$), otherwise $\Delta(ijkl)$ = 0. Thus helices Π are not stabilized. The parameter $\Delta'(ijkl)$ =1 if k and l satisfy either conditions: (k, l) = (i+2, $j-2$) or (i+2, j+2); otherwise $\Delta'(ijkl)$ = 0.

2.2 Sampling Tools

Three sampling tools can be used with the OPEP force field: ART,[9,10] MD[11] and REMD[12]. ART generates trajectories of several thousands of events and brings at each event the system from one relaxed state to another, going through an activation barrier (Fig.1B), and accepts or rejects the move according to the Metropolis criterion. By efficiently crossing energy barriers, ART therefore allows for a rapid sampling of low energy conformations. In contrast to ART, MD, which solves Newton's equations of motion, can explore protein dynamics and REMD, which runs in parallel a series of MD simulations (or replicas) at various temperatures and exchanges them periodically using the Metropolis criterion, can

explore protein thermodynamics. In the applications presented here, the integration time-step is 1 fs, T is controlled by the Berendsen's bath and the free energy is calculated using F = -RT log H(x,y), where x and y are two reaction coordinates, R is the gaz constant, and H(x,y) is the histogram of x and y.

2.3 Benchmarks

A number of systems have been used to train and validate the OPEP parameters. First, the analytic OPEP form is sufficiently rich to discriminate native from non-native protein structures for 29 targets.[8] Second, the applicability of OPEP in folding was recently revisited on the 60-residue B domain of protein A and we found that ART-OPEP simulations recovered the experimental three-helix bundle starting from random states, but also explained the observed shift to another PDB topology upon mutations.[13] Third, we verified that MD-OPEP describes qualitatively correctly the dynamics of proteins around their native states, although the implicit solvent and the coarse-grained nature of the side-chains modify the clock.[11] Finally, we checked that REM-OPEP reproduces the structural and thermodynamical properties of non-amyloid peptides such as the second β-hairpin from protein G and the 20-residue Trp-cage, starting from randomly chosen states.[12]

3 Free Energy Surfaces of Multimers

An important question in protein aggregation is to characterize the free energy surface of small multimers (e.g. dimers or heptamers) because the populated species may represent building blocks for further assembly.

We first probed the free energy surface of the Aβ(16-22) dimer, resulting from a 50 ns REMD-OPEP simulation starting from two disordered chains in random orientation. We used eight replicas with T varying between 287 and 500 K with exponential distribution and an exchange time between neighboring replicas of 2 ps, leading to an acceptance ratio between 30%-40%.

The free energy surface at 310 K projected on the cosine of the angle between the two KLVFFAE chains and the extended status of the chains, i.e. the product of the end-to-end distance of the chains divided by the product of the end-to-end distance for two ideal β-strands, shows multiple free energy minima. These are in-register and out-of-register parallel and antiparallel β-sheets, parallel loops and antiparallel loops, and cross conformations. It is interesting that all these states have been described by all-atom REMD simulations in explicit solvent.[14] Overall however, the dimer is found disordered, with a calculated random coil signal of 64% at 310 K.

Does the population of β-sheets change in higher-order species? To this end, we performed 50 MD simulations of 100-300 ns on β2m(83-89) heptamer at 310 K.[15] Starting from disordered states, we find that amorphous aggregates still represent 65% of the populated states, albeit well-ordered morphologies exist. These include the cross-β structure observed experimentally and shown in Fig.1C,[16] with Cα..Cα distances of 5.0 Å between the strands and 10.0 Å between the layers, and orthogonal β-sheets with only the meridional 5.0 Å reflection. Of particular interest is the finding of a variety of open and closed β-barrels consisting of six and seven chains with a Bolztmann probability on the order of 10%, i.e. a topology observed for KFFE hexamers[17] and Aβ(16-22) hexamers[4] using other simulation protocols.

4 From Random States to Fibrillar-Like Morphologies

Based on ART-OPEP and MD-OPEP simulations of KFFE,[17] NFGAIL[18] and β2m(83-89)[19] of various oligomeric sizes (from 6-mers to 16-mers), we could extract a generic aggregation picture of peptides with chain length less than 10 amino acids. Starting from random orientations of the chains and random coil conformations of each chain, the peptides first come together to form amorphous aggregates with a two-stranded or three-stranded β-sheet rapidly in place. Then, the system can remain trapped for a very long time in amorphous aggregates characterized by a total number of side-chain – side-chain and main-chain H-bonding interactions very similar those observed within well-ordered β-sheet structures. Or the system finds pathways to evolve very slowly to transient orthogonal and parallel β-sheets or a variety of closed and open β-barrels. Fig. 1D shows the closed β-barrel obtained from MD-OPEP simulations of 16 β2m(83-89) chains at 310 K and a concentration of 12mM. This β-barrel, which displays an inner diameter of 1.1 nm (minimal interior side-chain – side-chain distance) and is intriguing in terms of cytoxicity, can be formed from amorphous or orthogonal β-sheets.[15]

5 Concluding Remarks

We have described a coarse-grained protein simulation package to study folding and aggregation. The OPEP force field is generic and can be used to study the structural and thermodynamical properties of any amino acid sequence using MD and REMD, or explore the low energy conformations using ART. Based on ART-OPEP and MD-OPEP simulations, we could predict the importance of reptation moves of the chains in the final steps of aggregation, a mechanism later confirmed experimentally.[17] Application of this package to amyloid-forming peptide aggregation can complement experiments by pointing to the non-negligible Boltzmann probability of unexpected topologies, such as the β-barrel. This package is now used to study the aggregation of the Aβ(1-40)/Aβ(1-42) proteins associated with Alzheimer's disease and fibril disassembly caused by inhibitors, as reported in an accompanying paper.

Acknowledgments

I am indebted to CNRS, University of Paris 7 and "ImmunoPrion, FP6-Food-023144, 2006-2009" for financial support. Most of the work was done in collaboration with Normand Mousseau (University of Montreal, Canada), Guanghong Wei (Fudan University, China) and Yassmine Chebaro (CNRS, IBPC, Paris)

References

1. C. Dobson, *Protein folding and misfolding*, Nature **426**, 884–890, 2003.
2. S. Lesne, M.T. Koh, L. Kotilinek, R. Kayed, C.G. Glabe, A. Yang, M. Gallagher and K.H. Ashe, *A specific amyloid-beta protein assembly in the brain impairs memory*, Nature **440**, 352–7, 2006.

3. B. Ma and R. Nussinov, *Simulations as analytical tools to understand protein aggregation and predict amyloid conformation*, Curr Opin Chem Biol. **10**, 445–52, 2006.
4. A. Irback and S. Mitternacht, *Spontaneous beta-barrel formation: an all-atom Monte Carlo study of Abeta16-22 oligomerization*, Proteins **71**, 207–214, 2008.
5. J.H. Meinke and U.H. Hansmann, *Aggregation of beta-amyloid fragments*, J. Chem. Phys. **126**, 014706, 2007.
6. J. Khandogin and CL. 3rd Brooks, *Linking folding with aggregation in Alzheimer's beta-amyloid peptides*, Proc Natl Acad Sci U S A. **104**, 16880–5, 2007.
7. R. Pellerin, E. Guarnera and A. Caflisch, *Pathways and intermediates of amyloid fibril formation*, J Mol Biol. **374**, 917–24, 2007.
8. J. Maupetit, P. Tuffery and P. Derreumaux, *A coarse-grained protein force field for folding and structure prediction*, Proteins **69**, 394–408, 2007.
9. N. Mousseau, P. Derreumaux, G.T. Barkema and R. Malek, *Sampling activated mechanisms in proteins with the activation-relaxation technique*, J Mol Graph Model. **19**, 78–86, 2001.
10. S. Santini, N. Mousseau and P. Derreumaux, *In silico assembly of Alzheimer's Abeta16-22 peptide into beta-sheets*, J Am Chem Soc. **126**, 11509–16, 2004.
11. P. Derreumaux and N. Mousseau, *Coarse-grained protein molecular dynamics simulations*, J. Chem. Phys. **126**, 025101, 2007.
12. Y. Chebaro, X. Dong, R. Laghei, P. Derreumaux and N. Mousseau, *Replica exchange molecular dynamics simulations of coarse-grained proteins* , in preparation 2008.
13. J.F. Saint-Pierre, N. Mousseau and P. Derreumaux, *The complex folding pathways of protein A suggest a multiple-funnelled energy landscape*, J. Chem. Phys. **128**, 045101, 2008.
14. S. Gnanakaran, R. Nussinov and A.E. Garcia, *Atomic-level description of amyloid beta-dimer formation*, J Am Chem Soc. **128**, 2158–9, 2006.
15. W. Song, G.H. Wei, N. Mousseau and P. Derreumaux, *Self-Assembly of the beta2-Microglobulin NHVTLSQ Peptide Using a Coarse-Grained Protein Model Reveals a beta-Barrel Species*, J Phys Chem B. **112**, 4410–8, 2008.
16. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D. Structure of the cross-beta spine of amyloid-like fibrils. Nature 2005; 435:773-8.
17. N. Mousseau and P. Derreumaux, *Exploring the early steps of amyloid peptide aggregation by computers*, Acc. Chem. Res. **38**, 885–91, 2005.
18. A. Melquiond, J.C. Gelly, N. Mousseau and P. Derreumaux, *Probing amyloid fibril formation of the NFGAIL peptide by computer simulations*, J. Chem. Phys. **126**, 065101, 2007.
19. G.H. Wei, W. Song, P. Derreumaux and N. Mousseau, *Self-assembly of amyloid-forming peptides using molecular dynamics simulations and the OPEP coarse-grained force field*, Front Biosci. **13**, 5681–5692, 2008.

Milestoning

Ron Elber and Anthony West

Department of Chemistry and Biochemistry, ICES, C0200
The University of Texas at Austin, Austin TX 78712, USA
E-mail: ron@ices.utexas.edu

Milestoning is a theory and an algorithm to compute kinetics and thermodynamics of complex molecular systems. It makes it possible to study general processes on rugged energy landscapes on timescales not approachable by straightforward Molecular Dynamics (microseconds and milliseconds). The algorithm is based on monitoring progress along a set of discrete states (Milestones) using short-time microscopic trajectories that capture local dynamics. These discrete states can be (for example) hypersurfaces perpendicular to a reaction coordinate. Transition times between Milestones are recorded to produce Local-First-Passage-Time-Distributions (LFPTD). The theory is based on a non-Markovian integral equation for the probability flow between Milestones. The integral equation is equivalent to the Generalized Master equation. No specific model is assumed for the microscopic dynamics. The theory uses the LFPTD to compute the overall kinetic and thermodynamic. Complex transitions in proteins were investigated (allosteric transition in Scapharca hemoglobin, the recovery stroke in myosin). In the present review only the simple example of alanine dipeptide is discussed.

1 Introduction

Atomically detailed simulations provide useful information on biomolecular processes using a single unified model. Specifically, Molecular Dynamics (MD) algorithms are available to compute efficiently thermodynamic and equilibrium behavior. However, MD is limited when studying non-equilibrium processes and kinetics. Straightforward and typical trajectories of condensed phase systems rarely exceed hundreds of nanoseconds, far too short to investigate the kinetics of many interesting biophysical systems. Examples are of conformational transitions, ion permeation, protein folding and more. Extending the time scale of molecular simulations is therefore an important research direction and has attracted the attention of many investigators.

It is useful to classify processes of long time dynamics into two categories: Dynamics which are (i) activated or (ii) diffusive (figure 1). Significant progress has been made in algorithm design and theory development for activated processes[2, 14, 8, 6, 11]. In activated processes rare short time trajectories pass over significant free energy barriers and determine the overall kinetics. Progress has been slower for diffusive processes (or a mixture of activated and diffusive processes) in which the times of the individual transitional trajectories are intrinsically long. Diffusion on rugged energy landscapes is not necessarily associated with a narrow transition domain between stable states. A narrow transition domain is typical in activated processes and facilitates the use of short time trajectories to probe reactive events. If we probe an activated system at different time slices, in the majority of the observations we do not observe something new. The system remains in the reactant state until a rapid (but rare) transition is initiated to the product state. In contrast, probing diffusive processes show spatial progress in sequential observations. Milestoning is a theoretical and computational approach that aims at diffusive or mixed processes. Nevertheless, it can

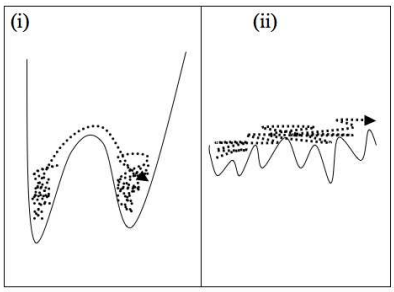


Figure 1. A schematic representation of an (i) activated and (ii) diffusive energy landscape that leads to different (corresponding) types of dynamics.

also handle activated processes and therefore suggests a uniform technology for the two types of dynamics.

A conceptual approach to long time dynamics is that of coarse graining in space and time. Indeed a number of groups have followed this idea, and have proposed fitting parameters of a kinetic model[13,1] or of the diffusion equation[16] based on atomically detailed simulations. For example, it is assumed that rate constants (exponential relaxations in time) describe transitions between the states of a Master equation. Power law and stretched exponential kinetics were found in biophysical kinetics[7]. Moreover, there is no rigorous mapping from an atomically detailed description of the system to a diffusion equation and the decision of what exactly to fit is not unique.

In contrast to the phenomenological modeling of the Master equation there is a rigorous approach to spatial and temporal coarse graining by Zwanzig and Mori. It is the Generalized Langevin Equation[17] or equivalently the Generalized Master Equation[10]. A memory kernel (and not a rate constant) describes the impact of the "bath". Unfortunately, the numerical calculations of the memory kernel of the Generalized Master Equation are difficult, motivating the use of the simpler and less rigorous Master Equation. At the core of the Milestoning approach one finds an algorithm to circumvent the difficulty in computing the memory kernel. The function we compute is formally equivalent but easier to estimate numerically than the rate kernel. Therefore the Milestoning approach is equivalent to the Generalized Master Equation and is based on a rigorous theory of non-equilibrium processes.

2 Milestoning

2.1 Theory

We describe below the approach of Milestoning. The Milestoning theory follows from an intuitive expression in which we capture the characteristics of the microscopic dynamics in a non-Markovian kernel, and then solve the dynamics within the assumption of local equilibrium.

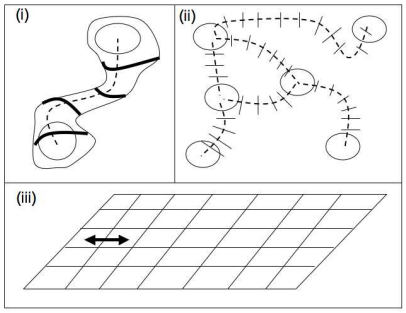


Figure 2. Examples of concrete realizations of the Milestoning idea: (i) Hypersurfaces (thick lines) perpendicular to a reaction coordinate (dashed line), (ii) hypersurfaces along reaction coordinates (thick dashed line) connecting local free energy minima between local free energy minima, and (iii) a grid over a few collective variables (illustrated for two). The Milestones are the boundaries of the squares. A potential transition at one interface it denoted with the double-ended arrow.

Let the vector describing the atomically detailed system be $X \in R^N$. We have at our disposal an operator D that generates trajectories such that $D\left[X\left(0\right),t\right] = X\left(t\right)$ where t is the time. Depending on the type of dynamics at hand X is a coordinate or a phase-space vector. The numerical application of D is usually made in small time steps and is repeated many times to produce a long time trajectory $X\left(t\right) = D^N\left[X\left(0\right),\Delta t\right], \Delta t = t/N$. The generation of the trajectories is by far the most expensive calculation at hand (also in Milestoning). Typically we will be interested in an ensemble of reactive trajectories that transition from a state of reactant to a state of product. The direct sampling of the transition (i.e. computing $X\left(t\right) \in$ product while $X\left(0\right) \in$ reactant) is assumed too expensive to pursue with direct use of D (of course, if it is not too expensive an exact solution of the problem is always better than a solution based on approximations or physical assumptions).

We now discuss the coarse grained model that we use to analyze the microscopic data and extend the time scale of the simulation. We partition the space and define discrete states s. Two of these states $s = 0$ and $s = L$ are the reactants and products respectively. Other $L-1$ states $s = 1, ..., L-1$ are called "Milestones". A realization of the space partition and "Milestones" is of (i) hypersurfaces perpendicular to a reaction coordinate, (ii) local free energy minima, and (iii) a grid over a reduced set of collective variables (figure 2). The examples discussed in the present manuscript (and in our published papers[6,4,15]) are of hypersurfaces perpendicular to reaction coordinates. However, the formulation below applies just as well to (ii) and (iii).

The probability of being at Milestone s at time t is $P_s\left(t\right)$. To compute $P_s\left(t\right)$ we derive from atomically detailed simulations the transition probability density $K_{s,s'}\left(\tau\right)$. It is the probability of making a transition from state s into state s' after an "incubation/waiting" time τ in the state s. The incubation time captures memory effects in which the transition probability between s and s' depends on the time spent already in s. It is the only microscopically derived function that we need for Milestoning. Note that we assume that the transition probability is independent of the absolute time. This assumption is not valid in systems that strongly deviate from equilibrium or a stationary state. We argued and illustrated[4,15] that it is much easier to compute the above matrix than to perform the complete simulations from reactants to products. Scaling arguments in support of the expected speed up are presented in the section Algorithm. This matrix is finally used in a probabilistic non-Markovian framework to obtain the overall kinetics of the system.

For further development it is convenient to define another function $Q_s\left(t\right)$. It is the

probability density that a trajectory will make a transition into s at time t. With $K_{s,s'}(\tau)$ and the definition of $Q_s(t)$ at hand, equation (1) below simply balances transition probabilities. The system is initiated at zero time and starting probabilities are injected into the Milestones. At later times we consider transitions between the states. To make a transition into s from one of the nearby s' it is necessary to transition first to s', wait (or incubate) at s' for time τ and then transition into s. The probability density of making a transition into s' at time $t-\tau$ is $Q_{s'}(t-\tau)$. The probability density of making a transition from s' to s after waiting time τ is $K_{s,s'}(\tau)$. Finally, a summation over all states s' that are directly connected to s, and over all incubation times τ gives equation (1) below.

$$Q_s(t) = \delta\left(t-0^+\right) P_s(0) + \int_0^t \sum_{s'} K_{s,s'}(\tau)\, Q_{s'}(t-\tau)\cdot d\tau \tag{1}$$

It is a matrix-vector equation in s space and an integral equation in time. The unknown is the vector of functions $Q_s(t)$. This equation is called in physics CTRW (Continuous Time Random Walk) and was used in phenomenological modeling of transport[9]. We are connecting this equation with atomically detailed simulations and use it to study long time phenomena. Microscopic dynamics is used to compute $K_{s,s'}(\tau)$.

Our interest focuses on $P_s(t)$, the probability of being at s at time t. With $Q_s(t)$ determined from equation (1) we write $P_s(t)$ as the integral of probabilities to make a transition into s at an earlier time t' and to remain at s (avoid transitions to other states s') until time t. Summation over all channels s' gives

$$P_s(t) = \int_0^t Q_s(t') \left[1 - \sum_{s'} \int_0^{t-t'} K_{s',s}(\tau)\cdot d\tau\right] dt' \tag{2}$$

Equations (1) and (2) are very general. They do not assume concr ete dynamics or mechanism, only that a transition matrix (with memory) can be defined (and computed) between states that represent the system. The solution of (1) and (2) is now a topic in applied mathematics and was obtained with different approaches[6,12,15]. The most obvious one is to solve the integral equation (equation (1)) by small time steps. Since the number of degrees of freedom was greatly reduced and the functions considered are much smoother with respect to time compared to MD, the computational efforts are still negligible compared to the calculations of the trajectories.

$$\begin{aligned} Q_s(0) &= P_s(0)/\Delta t \\ Q_s(\Delta t) &= Q_s(0) + \sum_{s'} K_{s,s'}(\Delta t)\cdot Q_{s'}(0)\cdot \Delta t \\ Q_s(2\Delta t) &= Q_s(0) + \sum_{s'} \left[K_{s,s'}(\Delta t)\cdot Q_{s'}(\Delta t) + K_{s,s'}(2\cdot\Delta t)\cdot Q_{s'}(0)\right] \\ &\ldots\ldots \end{aligned} \tag{3}$$

Another solution is based on Laplace transforms on (1) and (2) and algebraic manipulation of the transforms[12]. We quote only one result. Define the off diagonal matrix $\left(\hat{K}\right)_{s,s'}(t) = K_{s,s'}(t), s \neq s'$, the time integral $\langle f\rangle \equiv \int_0^\infty f(t)\, dt$ and an average over an ensemble of trajectories by $\bar{f}$. A useful measure for the kinetic properties of the system is the overall first passage time. It is defined as the time required for a trajectory initiated

at the reactant to reach the product state for the first time. The ensemble average of the overall first passage time $\bar{\tau}$ is given by

$$\bar{\tau} = I \cdot \left\langle \tau \cdot \hat{K}_{s,s'}(\tau) \right\rangle \left[I - \left\langle \hat{K}_{s,s'}(\tau) \right\rangle \right]^{-1} \cdot \varepsilon_i \tag{4}$$

where I is the identity matrix and ε_i is a unit vector in the direction of the initial Milestone i. The last Milestone is set to be absorbing.

2.2 Algorithm

It is important to emphasize that the theory described in the previous section is "equation free". We do not assume Langevin, Brownian or Newtonian mechanics. All reasonable forms of dynamics can be used to generate numerical values of the transition matrix. We compute the transition matrix as discussed below and we do not fit parameters to a particular coarse-grained dynamical model.

The only task of the microscopic dynamics is to compute the transition matrix $K_{s,s'}(\tau)$. This matrix is used in equation (1-4) to determine the overall rate and the evolution of the system in time. To facilitate the calculation and make it highly efficient compared to straightforward MD we restrict our attention to systems that satisfy the following requirements:
Condition (i): The system is in a stationary state (called ρ_s for Milestone s). Only a few variables may be left non-stationary.

Condition (ii): Trajectories that arrive at Milestone s' are distributed in the hyperplane according to the stationary distribution $\rho_{s'}$.

If the system is in equilibrium, or sufficiently close to it then assumption (i) is obviously satisfied. If the equilibrium is canonical then the probability density is $\rho_s \simeq \exp[-\beta U(X)]$, $X \in s$. This is the weight that was used in the alanine dipeptide example discussed in this paper. Alternatively, it is possible to have the system at a stationary (time-independent) but non-equilibrium state. An example is of a stationary flow of liquid in a confined environment. The second condition is subtler than the first. It requires the states s to be well separated in time so that the distribution of trajectories arriving to s' is the same as the stationary distribution of condition (i), i.e. $\rho_{s'}$. For example, this condition is satisfied if the time scale for the transition is longer that the time scale to reach local equilibrium at each of the Milestones. In practice it is possible to monitor and vary the transition time scales by placing the Milestones sufficiently apart with the second condition in mind. Of course, the choice of Milestoning should make the transition times between Milestones much shorter than the overall time scale of the reaction. Otherwise, Milestoning is not advantageous to MD.

The calculation of $K_{s,s'}(\tau)$ is done by sampling trajectories between a specific pair of states s and s'. The trajectories are initiated at s according to the stationary distribution ρ_s and their termination times at s' are recorded. These distributions are binned to estimate the probability that the system will transition between s and s' after incubation time τ. They are called the Local-First-Passage-Time-Distribution (LFPTD), and are also the matrix elements $K_{s,s'}(\tau)$. The LFPTD is normalized such that $f_{s,s'} = \int_0^\infty K_{s,s'}(\tau)\, d\tau$ is the

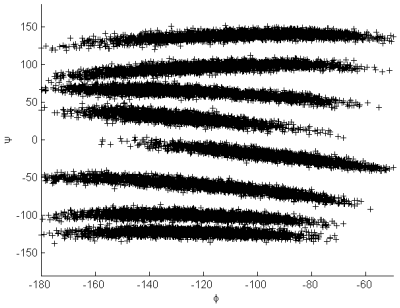


Figure 3. Sampling equilibrium distributions in the Milestones (hyperplanes perpendicular to the reaction coordinate) of the transition in alanine dipeptide. The distributions are projected onto a two-dimensional surface of internal coordinates (ϕ, ψ). Note that all simulation includes all coordinates of the dipeptide's atom and the coordinates of periodic box, solvating water molecules.

fraction of trajectories that were initiated at s and terminate at s' $\left(\sum_{s'} f_{s,s'} = 1\right)$.

What is the gain in studying the kinetic with Milestones instead of using straightforward MD? This important question was discussed extensively[15]. Practical gains were clearly illustrated in the examples of 4, 15. For completeness we quote an argument for computational speedup expected for diffusive processes[15]. The required computational resources are proportional to the number of force evaluations and therefore to the lengths of trajectories that reach the product state. Consider a reaction in which the system diffuses freely for a length L. The time to react using straightforward trajectories is proportional to L^2. In Milestoning we chop the complete length L to (say) N pieces. The time to diffuse through one piece is $(L/N)^2$. There are N pieces and therefore the time to destination in Milestoning is $(L/N)^2 \cdot N = L^2/N$. We obtain a speed up proportional to the number of Milestones used. Exponential speedup for systems with free energy barrier can be illustrated as well[15].

In the next paragraph we discuss a concrete example. We consider progress along a reaction coordinate measured by passing hypersurfaces (Milestones) orthonormal to it. In figure 2.ii we sketch a terminating trajectory between Milestones s and $s-1$. An ensemble of such trajectories is used to compute the probability densities $K_{s,s-1}(\tau)$ and the overall rate according to equation (4).

2.3 Example: Folding of a Solvated Dipeptide

We review a detailed calculation of the kinetics of a solvated dipeptide[15]. All the calculation described below were performed with the MOIL program which is in the public domain[5]. The Milestones were constructed from adiabatic energy surface of the ψ dihedral angle. The energy of the peptide in vacuum was minimized with the ϕ dihedral angle constrained to 150^o the ψ dihedral angle constrained to values between -180^o and +180^o degrees with a step size of 2.5 degrees. The minimizations provided a total of 144 structures (and potential Milestones) that we denote by $s = 1, ..., 144$. Milestones, which are hyperplanes perpendicular to the reaction coordinate, are defined by the coordinate of the minimized structures X_s, and the numerically estimated normals to the hyperplanes $q_s \equiv \frac{X_{s+1} - X_{s-1}}{|X_{s+1} - X_{s-1}|}$.

The different configurations were solvated in water boxes of volume (20Å)^3 and 248 water molecules. Molecular dynamics simulations constrained to each of the hyperplanes[3]

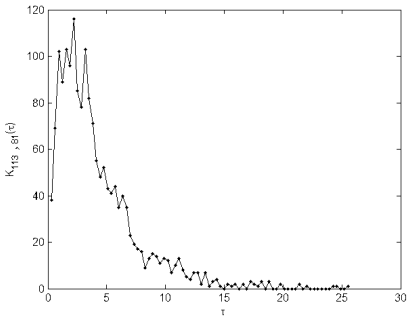


Figure 4. A typical example for a local first passage time distribution in a solvated alanine dipeptide. The example below is for a seven milestone partitioning of the reaction coordinate. The computed transition is from Milestone six to five (the indices below refer to the total of 144 Milestones that we considered using numerous variations[15]. The calculation with seven Milestones gives the correct rate. Even on this very simple example we obtain significant speedup. The average transition time between any two Milestones in the seven Milestone run is about 3.6 picoseconds. A cost of a single trajectory that passes all Milestones from the first to the last will be $3.6 \times 6 = 21.6$ picoseconds. This is about 3 times shorter than the correct first passage time of about 60 picoseconds.

were used to sample equilibrium distributions in the Milestones. In figure 3 we display a projection of the simulations on a (ϕ, ψ) map. Since the hyperplanes are defined in Cartesian space the projection on the space of internal coordinates show significant "width". We verified that the simulations create exact hyperplanes in Cartesian space.

M	$\bar{\tau}(ps)$	$\tilde{\tau}(fs)$
144	500(3.1)	31.2
74	261(1.2)	57.7
73	330(1.6)	58.3
37	104(0.63)	129
19	62(0.47)	373
11	53(0.50)	1,305
7	62(0.73)	3,581
5	68(0.93)	10,902
3	64(1.04)	-

Table 1. A summary of runs for alanine dipeptide conformational transition from an alpha helix to an extended chain conformation. The first column is the number of Milestones. The case of three Milestones (last row) is exact. The second column is the estimated overall first passage time and the third column the average LFPTD ($\tilde{\tau}$ - equation (5)). The runs differ in the number of Milestones used. When the number of Milestones is larger then the speedup is more significant. However if the number of Milestones is too large the local equilibrium assumption is violated and the rate is inaccurate. We compare the velocity decorrelation time (400 femtoseconds) with the average transition time between Milestones. If the number of Milestones is larger than 19 then the LFPTD is shorter than 400 femtoseconds and the rate is wrong. If the numbers of Milestones is smaller than 19 the results are quite accurate.

In the next step we initiate MD trajectories starting from the configurations sampled at each of the Milestones, s. The trajectories were integrated to termination at Milestones $s \pm 1$. The distributions of termination times are the elements of the transition matrix $K_{s,s\pm1}(t)$. A typical result is shown in figure 4. Initially we have used 144 Milestones. However, in this case the termination times are very short and they do not satisfy the condition about relaxation to equilibrium (condition (ii)). A useful test for relaxation is the velocity decorrelation time that we found in that case to be about 400 femtoseconds (figure

8 of 15). A typical transition time between Milestones, estimated as the following average

$$\tilde{\tau} \equiv \frac{1}{L}\sum_{s}\bar{\tau}_s = \frac{1}{L}\int_0^{\infty}\tau_s \cdot \left(K_{s,s+1}\left(\tau\right) + K_{s,s-1}\left(\tau\right)\right)d\tau \tag{5}$$

The typical transition time must be larger than the velocity de-correlation time. In table 1 we show that this condition is satisfied only for a number of milestones smaller than 19. All the calculations with a number of milestones smaller or equal to 19 approximate well the exact rate.

This calculation illustrates that Milestoning provides accurate results for a non-trivial but exactly solvable model. It was also shown that it is significantly more efficient. Comparing directly simulation time, we have concluded in our publication that in this example Milestoning was more efficient by a factor of about 9. In other calculations of larger more complex systems (e.g. the allosteric transition in Scapharca hemoglobin), the speedup was a factor of about 1,000.

References

1. Chodera, J. D., W. C. Swope, et al. (2006). *Long-time protein folding dynamics from short-time molecular dynamics simulations.* Multiscale Modeling & Simulation **5**(4): 1214-1226.
2. Dellago, C., P. G. Bolhuis, et al. (2002). *Transition path sampling.* Advances in Chemical Physics **123**: 1-78.
3. Elber, R. (1990). *Calculation of the potential of mean force using molecular dynamics with linear constraints – An application to a conformational transition in a solvated dipeptide.* Journal of Chemical Physics **93**(6): 4312-4321.
4. Elber, R. (2007). *A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin.* Biophysical Journal **92**(9): L85-L87.
5. Elber, R., A. Roitberg, et al. (1995). *Moil a program for simulations of macromolecules.* Computer Physics Communications **91**(1-3): 159-189.
6. Faradjian, A. K. and R. Elber (2004). *Computing time scales from reaction coordinates by milestoning.* Journal of Chemical Physics **120**(23): 10880-10889.
7. Frauenfelder, H., B. H. McMahon, et al. (2001). *The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin.* Proceedings of the National Academy of Sciences of the United States of America **98**(5): 2370-2374.
8. Hummer, G. and I. G. Kevrekidis (2003). *Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations.* Journal of Chemical Physics **118**(23): 10762-10773.
9. Kenkre, V. M. and R. S. Knox (1974). Generalized Master Equation Theory of Excitation Transfer.
10. Mori, H., H. Fujisaka, et al. (1974). A new expansion of the master equation.
11. Ren, W., E. Vanden-Eijnden, et al. (2005). *Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide.* Journal of Chemical Physics **123**(13).

12. Shalloway, D. and A. K. Faradjian (2006). *Efficient computation of the first passage time distribution of the generalized master equation by steady-state relaxation.* Journal of Chemical Physics **124**(5).
13. Sriraman, S., L. G. Kevrekidis, et al. (2005). *Coarse master equation from Bayesian analysis of replica molecular dynamics simulations.* Journal of Physical Chemistry B **109**(14): 6479-6484.
14. Voter, A. F., F. Montalenti, et al. (2002). *Extending the time scale in atomistic simulation of materials.* Annual Review of Materials Research **32**: 321-346.
15. West, A. M. A., R. Elber, et al. (2007). *Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide.* Journal of Chemical Physics **126**(14).
16. Yang, S. C., J. N. Onuchic, et al. (2007). *Folding time predictions from all-atom replica exchange simulations.* Journal of Molecular Biology **372**(3): 756-763.
17. Zwanzig. R. (2001). Nonequilibrium statistical mechanics. Oxford, Oxford University Press.

Simulating Biomolecules in Cellular Environments

Michael Feig[1,2], **Seiichiro Tanizaki**[1,5], **Jana Chocholoušová**[1,6], **Maryam Sayadi**[2], **Jacob W. Clifford**[3], **Brian D. Connelly**[4], **Shayantani Mukherjee**[1], **and Sean M. Law**[1]

[1] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
E-mail: {feig, shayan, slaw}@msu.edu

[2] Department of Chemistry, Michigan State University
E-mail: sayadima@msu.edu

[3] Department of Physics, Michigan State University
E-mail: jacobc@msu.edu

[4] Department of Computer Science and Engineering, Michigan State University
E-mail: connel42@cse.msu.edu

[5] *current address*: Department of Chemistry and Biochemistry, University of Texas, Arlington, Arlington, TX 76019, USA
E-mail: tanizaki@uta.edu

[6] *current address*: Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague 6, Czech Republic
E-mail: chocholousova@uochb.cas.cz

Simulations of biomolecules with realistic representations of cellular environments remain challenging. Implicit solvent methods can reduce the system degrees of freedom and accelerate conformational sampling. Implicit solvent methods can be developed based on a decomposition of the solvation free energy into electrostatic and non-polar contribution. The electrostatic contribution based on continuum electrostatics theory can be conveniently calculated according to the generalized Born (GB) formalism. While the GB formalism is well established in aqueous solvent, applications to dense cellular environments and heterogeneous biological membrane environments are discussed.

1 Introduction

Biological function is often not fully understood until a picture of biomolecular structure and dynamics at the atomic level is developed. Structures of proteins and nucleic acids have become widely available from X-ray crystallography and NMR spectroscopy while atomic-level insight into single molecule dynamics has been gained primarily from computational molecular dynamics studies. Molecular dynamics simulations are well established for the study of single biomolecules over nanosecond time scales, but it remains challenging to study larger biomolecular complexes and longer, biologically more relevant time scales[1].

A significant part of the computational cost of molecular dynamics simulations stems from solvent-solvent and solute-solvent interactions. In the canonical approach the solvent environments is represented in an explicit fashion often resulting in systems with many more solvent atoms than solute atoms. One approach for accelerating simulations of biomolecules involves the application of mean-field descriptions of solute-solvent interactions instead of explicit solvent[2]. A common strategy is the decomposition of the solvation

free energy into polar and non-polar components according to equation 1:

$$\Delta G_{\text{solvation}} = \Delta G_{\text{solvation,polar}} + \Delta G_{\text{solvation,non-polar}} \quad (1)$$

The polar comntribution of the solvation free energy is due to electrostatic solute-solvent interactions and can be obtained by invoking continuum electrostatic theory[3]. In this formalism, a solvated biomolecular system may be described as a set of explicit solute charges embedded in a low-dielectric cavity that is surrounded by a continuum high-dielectric environment. Such a system is described rigorously by the Poisson equation 2, which relates the electrostatic potential ϕ to a distribution of charges, ρ, and dielectric constants ϵ.

$$\nabla \cdot [\epsilon(r) \nabla \phi(r)] = -4\pi\rho(r) \quad (2)$$

It turns out that direct solution of the Poisson equation with finite difference methods is also computationally expensive. Instead, the empirical generalized Born (GB) formalism is more commonly employed to approximate the electrostatic solvation free energy from Poisson theory at a fraction of the computational cost[4].

Many flavors of the GB formalism have been proposed over the last two decades. All are based on the GB equation 3 proposed orignally by Still et al.[5]:

$$\Delta G_{\text{solvation,GB}} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-r_{ij}^2 / F \alpha_i \alpha_j}}} \quad (3)$$

Different GB implementations vary in the calculation of the GB radii α_i. Acoording to the Coulomb field approximation, the α_i are essentially obtained from an integral of $1/r^4$ over the solute cavity[6]. The most accurate GB methods estimate the integral directly and include additional correction terms to account for deficiencies of the Coulomb field approximation in larger molecules[7].

The non-polar contribution to the solvation free energy consists of the cost of solute cavity formation and contributions from solute-solvent van der Waals interactions. These two terms may be considered separately but are often combined into a single simple term based on the solvent-accessible surface area (SASA) according to Eq. 4[8] with typical values of γ ranging from $5 - 30$ cal/mol/Å^2.[8,9]

$$\Delta G_{\text{solvation, non-polar}} = \gamma \cdot SASA \quad (4)$$

2 Simulations of Biomolecules in Implicit Aqueous Solvent

Simulations of biomolecules in implicit solvent become possible by simply adding the solvation free energy to a vacuum molecular mechanics potential with the partial atomic charges used in the calculation of the electrostatic solvation free energy taken from the force field according to Eq. 5. The resulting forces may then be integrated according to Langevin dynamics[10], Eq. 6, to obtain trajectories that are coupled to solvent through stochastic collisions R and frictional forces according to the friction coefficient γ.

$$U_{implicit} = U_{MM} + \Delta G_{\text{solvation,GB}} + \gamma \cdot \text{SASA} \quad (5)$$

$$F = -\nabla U_{implicit} + R - \gamma \dot{r}(t) = m\ddot{r}(t) \quad (6)$$

Implicit solvent simulations represent a compromise between computational efficiency and the level of realism that can be achieved with a mean-field solvent representation. Model

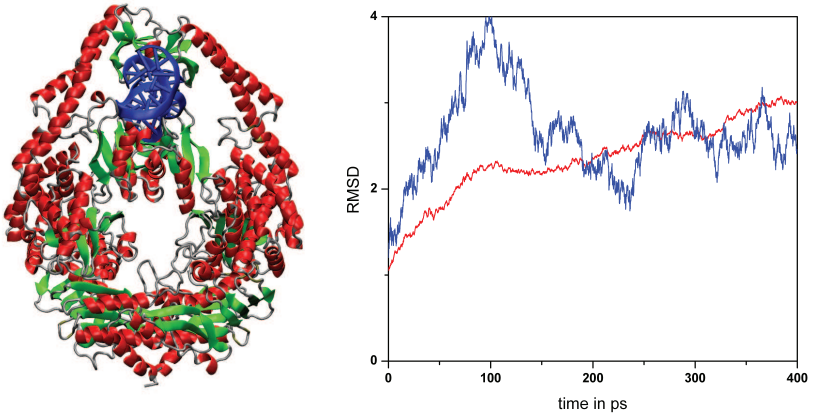


Figure 1. Left: Crystal structure of MutS in complex with DNA (PDB ID: 1E3M). Right: RMSD of C_α atoms (red) and DNA phosphorous atoms (blue) from the experimental structure in molecular dynamics simulations with implicit solvent using the GBMV method in CHARMM[7].

deficiencies may materialize due to the continuum nature of the implicit solvent model and simplicity of the non-polar term or due to limitiations in approximating the continuum electrostatic model with a particular GB implementation.

The quality of implicit solvent simulations may be evaluated by comparing long molecular dynamics simulations with implicit and explicit solvent with experimental data[11]. Very similar conformational sampling is observed in long simulations of protein G and ubiquitin over tens of nanoseconds. Average root mean square deviations (RMSD) from the experimental structures from X-ray crystallography of less than 1 Å with both implicit and explicit solvent[11].

Stable implicit solvent simulations of larger complexes and of nucleic acids[13] are also possible. As an example, simulations of the MutS dimer in complex with mismatched DNA is shown in Figure 1. The simulations are relatively short but reach only 3 Å after 400 ps which is considered good for such a large complex with flexible domains. Explicit solvent simulations of MutS reached a similar deviation from the X-ray structure within the first few hundred picoseconds (data not shown). Successful implicit solvent simulations of nucleic acid systems are remarkable because of strong solute-solvent interactions with the poly-ionic nucleic acids.

It turns out that the GBMV implicit solvent method used here does not offer any computational advantage on a time per integration step basis for the large MutS system because the ratio of solvent to solute atoms decreases with increasing solute size for a single, approximately spherical solute molecule. Previous timing tests have found that implicit solvent simulations with the relatively expensive but accurate GBMV method are only faster for single proteins with up to 200-300 residues[14]. However, implicit solvent offers additional advantages, in particular the ability to traverse conformational space more rapidly when using Langevin dynamics with low friction coefficients[11]. Furthermore, implicit solvent is the only practical solution for simulations of multiple freely diffusing biomolecules within a given solvent environment.

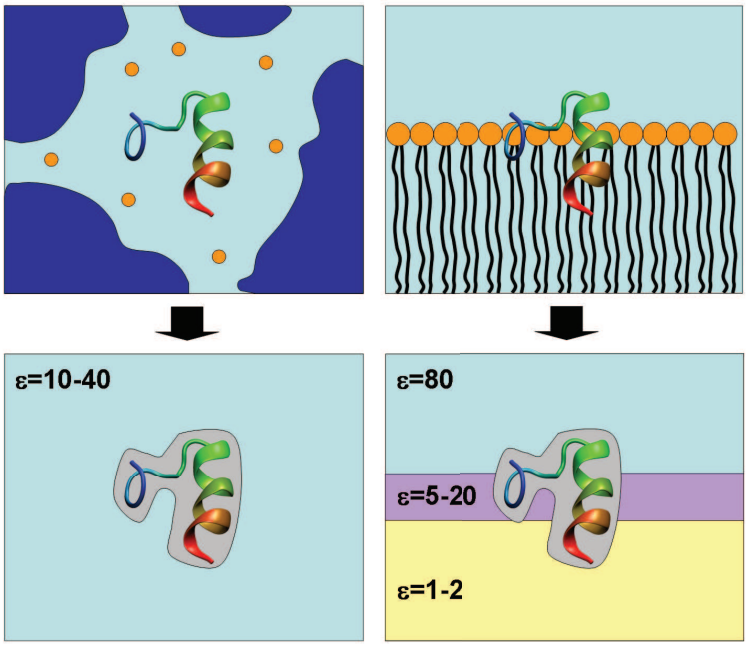


Figure 2. Schematic illustration of implicit modeling of dense cellular environments (left) and heterogeneous membrane bilayers (right) with continuum electrostatics.

3 Simulations of Biomolecules in Implicit Cellular Environments

Dilute aqueous solvent is generally not a good model of crowded cellular environments[15]. The dense concentration of biomolecules and co-solvents presents a complex solvent environment with reduced polarizability compared to water and steric hindrance due to crowding. As a first approximation, dense cellular environments can be modeled in an implicit fashion by assuming a reduced dielectric constant[17] and an increased cost of cavity formation to reflect crowding effects (see Fig. 2). This raises the question how the conformational sampling of peptides and proteins varies in environments with reduced dielectric response. Based on typical dielectric constants of proteins, the effects of co-solvents, and a dielectric modulation of water[16] in crowded environments, one may estimate that the effective dielectric constant of dense cellular environments lies in the range of $\epsilon = 10 - 40$.

Continuum electrostatics methods can readily accommodate a reduced dielectric constant of the environment. Based on physical insight it is expected that reduced dielectric screening enhances charge-charge interactions and in particular the formation of secondary structure elements through hydrogen bonding while the formation of hydrophobic cores becomes less favorable in low dielectric environments. Implicit solvent simulations with dielectric constants between 5 and 80 using a slightly modified GB formalism have confirmed these assumptions for poly-alanine and the amphipathic peptide melittin. However, the simulation results also indicate that even relatively minor changes in the dielectric constant can affect the conformational sampling of melittin in more subtle ways[17].

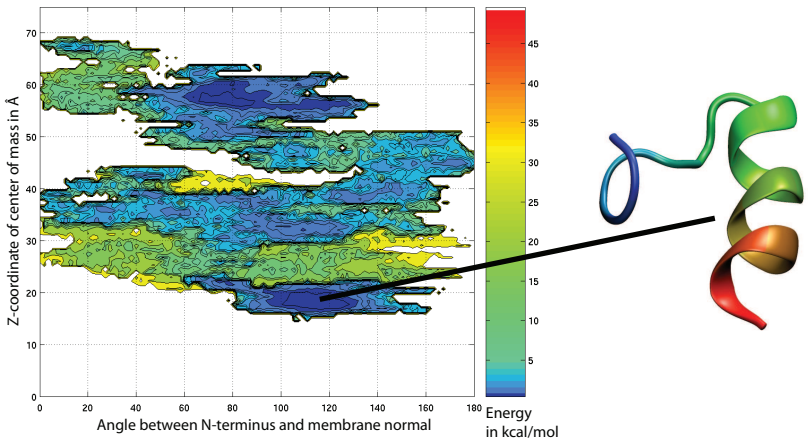


Figure 3. Conformational sampling of influenza fusion peptide with neutral N- and C-termini from temperature replica exchange simulations with the HDGB implicit membrane model. Each of the eight replicas, spaced from 300 to 500K, was run for 15 ns. The potential of mean force in kcal/mol is shown as a function of the angle of the N-terminal part of the peptide relative to the membrane normal and the z-coordinate of the center of mass in Å with z=0 corresponding to the membrane center. The dominant conformation is shown on the right.

4 Simulations of Biomolecules in Implicit Membrane Environments

Biological membrane environments involve heterogeneous environments with a hydrophobic core and a polar surrounding aqueous solvent environment. Implicit membrane models therefore require a spatially varying electrostatic and non-polar contribution to the solvation free energy. The electrostatic solvation free energy can be calculated based on a layered dielectric system[18] with a low dielectric core with ϵ near 1, an intermediate dielectric region near the membrane-water interface, and a high delectric environment elsewhere. Such a model is readily implemented with Poisson theory but presents challenges for the standard GB formalism. The application of heterogeneous dielectric environments becomes possible after introduction of an effective dielectric profile $\epsilon(z)$ and application of the modified GB equation 7:

$$\Delta G_{\text{solvation,HDGB}} = -\frac{1}{2}\sum_{i,j}\left(1-\frac{1}{(\epsilon_i+\epsilon_j)/2}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2+\alpha_i\alpha_j e^{-r_{ij}^2/F\alpha_i\alpha_j}}} \qquad (7)$$

The non-polar contribution to the solvation free energy also varies between an essentially zero cost in the membrane interior to a significant cost of cavity formation in water.

Using this scheme, it has been possible to perform implicit solvent simulations of membrane-bound peptides and proteins[18,19] and accurately reproduce explicit solvent membrane insertion profiles of amino acid side chain analogs[18]. As an example of first applications, Fig. 3 shows extensive conformational sampling of influenza fusion peptide near a membrane interface.

5 Summary and Outlook

Implicit solvent formalisms can be used successfully to simulate biomolecules in cellular environments ranging from simple aqueous solvent to dense cellular environments and heterogeneous dielectric environments. These methods open the door for the simulation of sub-cellular processes in atomic detail and over biologically relevant time scales.

Acknowledgments

Financial support from NSF CAREER grant 0447799 and the Alfred P. Sloan Foundation and access to computational resources through the High Performance Computing Center at Michigan State University are acknowledged.

References

1. M. Karplus, *Molecular dynamics simulations of biomolecules*, Acc. Chem. Res. **35**, 321–323, 2002.
2. B. Roux, T. Simonson, *Implicit Solvent Models* Biophys. Chem. **78**, 1–20, 1999.
3. B. Honig, A. Nicholls, *Classical Electrostatics in Biology and Chemistry*, Science **268**, 1144–1149, 1995.
4. M. Feig, C. L. Brooks, III, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, Curr. Op. Struct. Biol. **14**, 217–224, 2004.
5. W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*, J. Amer. Chem. Soc. **112**, 6127–6129, 1990.
6. D. Bashford, D. A. Case, *Generalized Born models of macromolecular solvation effects*, Ann. Rev. Phys. Chem. **51**, 129–152, 2000.
7. M. S. Lee, F. R. Salsbury, Jr., C. L. Brooks, III, *Novel generalized Born methods* J. Chem. Phys. **116**, 10606–10614, 2002.
8. D. Sitkoff, K. A. Sharp, B. Honig, *Accurate calculation of hydration free-energies using macroscopic solvent models*, J. Phys. Chem. **98**, 1978–1988, 1994.
9. W. Im, M. S. Lee, C. L. Brooks, III, *Generalized Born model with a simple smoothing function*, J. Comp. Chem. **24**, 1691–1702, 2003.
10. C. L. Brooks, III, M. Berkowitz, S. A. Adelman, *Generalized Langevin theory for many-body problems in chemical-dynamics - gas, surface collisions, vibrational-energy relaxation in solides, and recombination reactions in liquids*, J. Chem. Phys. **73**, 4353–4364, 1980.
11. M. Feig, *Kinetics from implicit solvent simulations of biomolecules as a function of viscosity*, J. Chem. Theory Comput. **3**, 1734–1748, 2007.
12. T. A. Darden, D. York, L. G. Pedersen, *Particle mesh Ewald: An N log(N) method for Ewald sums in large systems*, J. Chem. Phys. **98**, 10089–10092, 1993.
13. J. Chocholousova, M. Feig, *Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs. experiment*, J. Phys. Chem. B **110**, 17240–17251, 2006.

14. M. Feig, J. Chocholousova, S. Tanizaki, *Extending the horizon towards the efficient modeling of large biomolecular complexes in atomic detail*, Theor. Chem. Acc. **116**, 194–205, 2006.
15. A. B. Fulton, *How crowded is the cytoplasm?*, Cell **30**, 345–347, 1982.
16. F. Despa, A. Fernandez, R. S. Berry, *Dielectric modulation of biological water*, Phys. Rev. Lett., **93**, 2002.
17. S. Tanizaki, J. W. Clifford, B. D. Connelly, M. Feig, *Conformational sampling of peptides in cellular environments*, Biophys. J. **94**, 747–759, 2008.
18. S. Tanizaki, M. Feig, *A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes*, J. Chem. Phys. **122**, 124706, 2005.
19. S. Tanizaki, M. Feig, *Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model*, J. Phys. Chem. B **110**, 548–556, 2006.

Scalable Systems for Computational Biology

Christoph Pospiech

IBM Deutschland GmbH, Freiberger Straße 35, 01067 Dresden, Germany
E-mail: pospiech@de.ibm.com

Driving up processor clock speed is about to hit several limitations, a soaring energy consumption being among them. The BlueGene system® from IBM is shown to provide performance in a scalable and energy efficient way. Results from NAMD and similar codes show, how the system can be used in the area of Computational Biology.

1 The Promise of Moore's Law

The frequently quoted term "Moore's Law" goes back to a paper published in the Electronics Magazine[7]. In this paper, G. E. Moore states that

> "the complexity for minimum component costs has increased at a rate of roughly a factor of two per year ... Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer."

The "factor of two per year" was later relaxed into a "doubling every two years". This statement became known as "Moore's Law". It was quoted many times, not always in full length or precise form. So it started shifting its meaning. As a consequence, the wikipedia[11] entry for Moore's Law lists no less than 9 different formulations on exponential growth laws related to integrated circuits.

Hans Werner Meuer in a recent paper[6] uses linear regression on a logarithmic scale to fit data from the Top500 list to an exponential growth law. This is illustrated in Fig. 1. He finds a similar growth rate as "predicted" by Gordon Moore. It is therefore tempting to see this as another variation of Moore's law.

These data might generate the impression that computers grow with exponential rate in about every aspect. But there are already some limitations visible on the horizon. The original quotation of Moore's law was a statement about the number of components on an integrated circuit. Cramming more components to the same space forces the components to get smaller in size. In the current CMOS technology the gates sizes are now approaching molecular levels[9,8], and start leaking like a sieve. To make up for the lost electrons, the energy density has to be increased. In conjunction with an increased clock speed this leads to the heat flux curves as shown in Fig. 2[9]. The heat flux is directly connected to the energy consumption as energy is needed to generate the heat and also to cool it again.

It might be tempting to anticipate a new curve to the right of the existing two curves in Fig. 2, but there is yet no technology that allows jumping again to a cooler regime. One way to evade from the heat and energy trap is to trade low energy for the single core

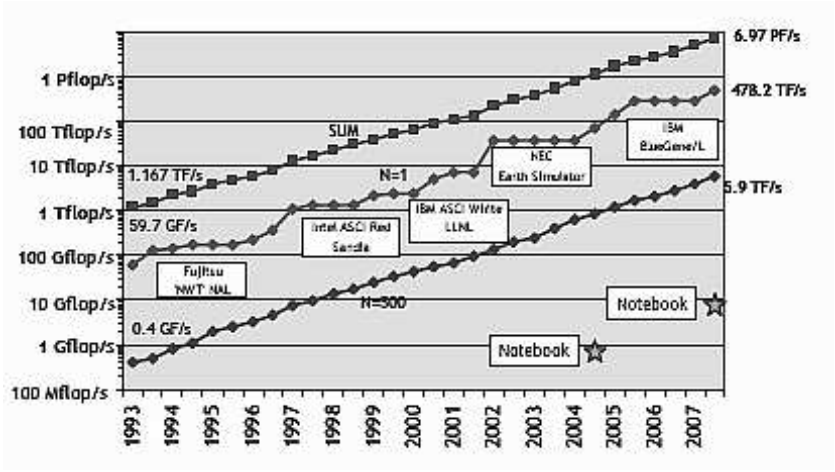


Figure 1. Top500 performance data fitted with linear regression on a logarithmic scale to an exponential growth law.

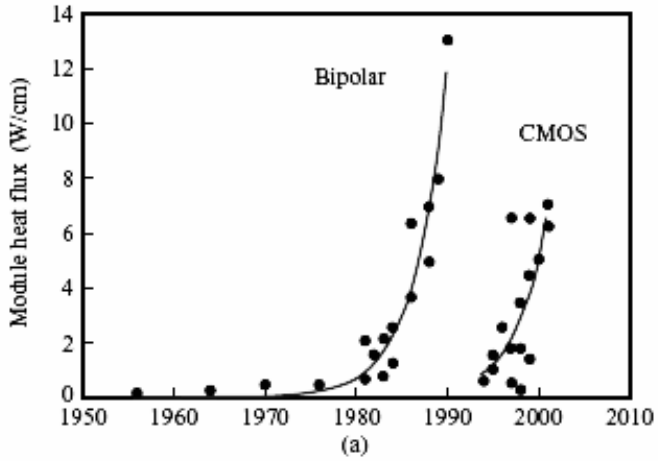


Figure 2. Module heat flux from various systems in bipolar and CMOS technology.

against high clock rate. To obtain high application performance out of low clocked cores, many cores have to be thrown at the problem. The IBM Blue Gene® is one example for a system that follows this concept. The "green500" list[10] from February 2008 shows that this concept may lead to systems with both, high performance and high energy efficiency.

2 Blue Gene

On June 26, 2007, IBM announced the Blue Gene/P™ system as the leading offering in its massively parallel Blue Gene® supercomputer line, succeeding the Blue Gene/L™ system. The following description of the Blue Gene/P™ system is based on a recent paper by the IBM Blue Gene® team[2].

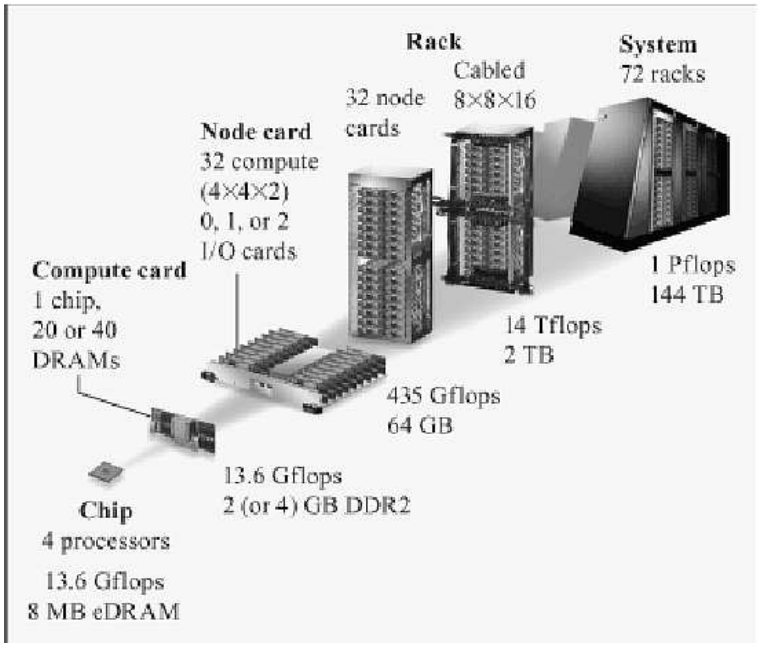


Figure 3. The packaging hierarchy of a Blue Gene/P™ system.

The packaging of the Blue Gene/P™ system is shown in Fig. 3. The main building block is a single ASIC (application-specific integrated circuit), with four IBM PowerPC 450 (PPC450)-embedded 32-bit processor cores, arranged as an SMP. A dual-pipeline floating-point unit (FPU) is attached to each PPC450 core. The design of this dual FPU is logically identical to the one used in the Blue Gene/L™ system. It supports two simultaneous double-precision floating-point calculations in SIMD (single-instruction, multiple-data) fashion, along with instruction set extensions for complex number arithmetic. The dual-pipeline FPUs can simultaneously execute two fused multiplyadd instructions per machine cycle, each of which is counted as 2 FLOPs (floating-point operations). Thus, each processor unit (PPC450 and FPU) has a peak performance of 4 FLOPs per machine cycle.

In addition to the compute nodes, the Blue Gene/P™ system contains a configurable number of I/O nodes. The I/O nodes are physically the same compute cards as described above, but their position in the system differentiates their logical function. I/O nodes have the 10-Gigabit Ethernet (GbE) interface enabled for communication with a file system and host computers.

Thirty-two compute cards and, optionally, up to two I/O cards are packaged onto the next-level board, called the node card. Sixteen node cards are plugged from both sides into a vertical midplane card, completing an assembly of 512 compute nodes in an $8 \times 8 \times 8$ configuration. The inbound and outbound network connections for this 512-way cube are routed to four link cards that carry a total of 24 Blue Gene/P™ link (BPL) chips. The assembly of 16 node cards, 4 link cards, and an additional service card is called a midplane or a 512-way. The BPL chips are relatively simple switches that, depending on the size and configuration of a user partition of the system, route network signals back into the midplane (completing the wraparounds for an $8 \times 8 \times 8$ torus) or route the network signals through cables to another midplane for larger partitions. Two midplanes, one on top of the other, complete a rack. Thus, a rack has 1,024 nodes, or 4,096 cores, giving a peak performance of 13.9 teraflops (Tflops). Scaling upward, a 72-rack system can package 72K nodes (288K cores) (where K stands for 1,024) into a 1-petaflops (Pflops) (peak) system, and larger configurations up to 256 racks (3.56 Pflops peak) are possible.

In the Blue Gene/P™ system, three networks are used for node-to-node communication: a 3D torus network, a collective network, and a global barrier network. On the Blue Gene/L™ system, the processor cores were responsible for injecting (or receiving) packets to (or from) the network. On the Blue Gene/P™ system, a direct memory access (DMA) engine has been added to offload most of this responsibility from the cores, thus enabling better overlap of communication and computation. Specifically, the DMA interfaces with the torus network. The combination of the DMA, torus network, and memory system is capable of supporting high-bandwidth communications.

The Blue Gene® system software approach is to start with a minimal-functionality system stack, but one that is designed to scale. The Blue Gene® strategy to achieve scalability and high performance has been to start simply, for example, with space sharing, one job per partition, no paging, and one thread per core.

3 Application Examples

3.1 CPMD (Car-Parrinello Molecular Dynamics)

The CPMD code is a plane wave/pseudopotential implementation of Density Functional Theory, particularly designed for ab-initio molecular dynamics. Its first version was developed by Jurg Hutter at IBM Zurich Research Laboratory starting from the original Car-Parrinello codes[1]. CPMD runs on many different computer architectures and it is well parallelized. The application is mainly written in Fortran, parallelized for distributed-memory with MPI, with an option to add an additional level of parallelism using OpenMP for multi-processor nodes. CPMD makes extensive use of three-dimensional FFTs, which require efficient all-to-all communication[3]. The scalability was improved using a task-group implementation of the FFT with a special mapping to the Blue Gene® torus network[4]. Moreover, overlap matrices, which were replicated in the standard CPMD code, have been distributed on a subset of the nodes to be able to handle large systems (more than 3000 electronic states). The single processor performance of CPMD was optimized for Blue Gene® using SIMD-enabled routines for the most common calls such as DGEMM, DCOPY, AZZERO, and FFT.

The following comparison was taken from benchmarks with a methanol vapor/liquid

OpenMP speedup	
2 OpenMP threads	4 OpenMP threads
1.91	3.6

Table 1. OpenMP speedup on Blue Gene/P^{TM} for CPMD.

interface ($\sim$700 atoms , 70Ry). Table 1 shows that the Blue Gene/P^{TM} takes significant advantage of an MPI/OpenMP hybrid programming paradigm.

3.2 NAMD (NAnoscale Molecular Dynamics)

NAMD (nanoscale molecular dynamics) is a production molecular dynamics (MD) application for biomolecular simulations that include assemblages of proteins, cell membranes, and water molecules. In a biomolecular simulation, the problem size is fixed and a large number of iterations must be executed in order to understand interesting biological phenomena. Hence, we need MD applications to scale to thousands of processors, even though the individual timestep on one processor is quite small. NAMD has demonstrated its performance on several parallel computer architectures.

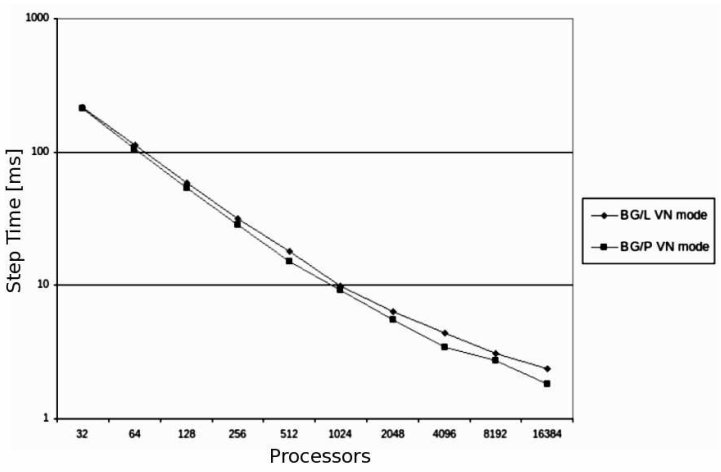


Figure 4. Comparison of NAMD execution times on Blue Gene/L^{TM} and Blue Gene/P^{TM}.

Figure Fig. 4 shows the comparison of times per execution step for Blue Gene/L^{TM} and Blue Gene/P^{TM}. The application was a 92K atom APoA1 benchmark with PME every 4 steps. The data for Blue Gene/L^{TM} were taken from a recent paper by S. Kumar et al.[5]. The Blue Gene/P^{TM} results are very recent and conveyed to the author via private communication.

The comparison shows that Blue Gene/P^{TM} performs better for all compared number of processors. In opinion of S. Kumar, the clock-speed only accounts for a 5% speed up. The DMA accounts for the rest of the difference.

4 Summary

Moore's law might generate the impression that computers grow with exponential rate in about every aspect, but there are already some limitations visible on the horizon. One way to evade from the heat and energy trap is to trade low energy for the single core against high clock rate. To obtain high application performance out of low clocked cores, many cores have to be thrown at the problem. The IBM Blue Gene® is one example for a system that follows this concept. The BlueGene system® from IBM is shown to provide performance in a scalable and energy efficient way. Results from NAMD and similar codes show, how the system can be used in the area of Computational Biology.

References

1. R. Car and M. Parrinello, *Unified Approach For Molecular Dynamics and Density Functional Theory*, Phys. Rev. Lett. **55(22)**, 2471, 1985.
2. IBM Blue Gene team, *Overview of the IBM Blue Gene/P project*, IBM Journal on Research and Development **52 no. 1/2**, 199–220, 2008.
3. J. Hutter and A. Curioni, *Dual-level parallelism for ab initio molecular dynamics: Reaching teraflop performance with the CPMD code*, Parallel Computing **31**, 1–17, 2005.
4. J. Hutter and A. Curioni, *Car-Parrinello Molecular Dynamics on Massively Parallel Computers*, ChemPhysChem **6**, 1788–1793, 2005.
5. S. Kumar et. al., *Scalable molecular dynamics with NAMD on the IBM Blue Gene/L system*, IBM Journal on Research and Development **52 no. 1/2**, 145–158, 2008.
6. H. W. Meuer, *The TOP500 Project: Looking back over 15 years of Supercomputing Experience*, to be published in INFORMATIK SPEKTRUM, Springer Verlag, `http://www.top500.org/files/TOP500_Looking_back_HWM.pdf`.
7. G. E. Moore, *Cramming more components onto integrated circuits*, Electronics Magazine **38**, no. 8, 1965, `ftp://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf`.
8. E. J. Nowak, *Maintaining the benefits of CMOS scaling when scaling bogs down* IBM Journal on Research and Development **46 no. 2/3**, 169–180, 2002.
9. R. R. Schmidt, B. D. Notohardjono, *High-end server low-temperature cooling*, IBM Journal on Research and Development **46 no. 6**, 739–751, 2002.
10. `http://www.green500.org/lists/2008/02/green500.php`.
11. `http://en.wikipedia.org/wiki/Moore's_law`.

High Performance Computing in Multiscale Modeling Cardiac Contraction: Bridging Proteins to Cells to Whole Heart

John J. Rice

IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA
E-mail: johnrice@us.ibm.com

The availability of increased computing with thousands of computational cores enables new classes of biological models that include detailed representations of proteins and protein complexes with spatial interactions. We develop such a model of the interaction of actin and myosin in the cardiac sarcomere. The model includes explicit representations of actin, myosin, and regulatory proteins. Although this is not an atomic-scale model, as would be the case for molecular dynamics simulations, the model seeks to represent spatial interactions between protein complexes that are thought to produce characteristic cardiac muscle responses at larger scales. While the model simulates the microscopic scale, when model results are extrapolated to larger structures, the model recapitulates complex, nonlinear behavior such as the steep calcium sensitivity of developed force in muscle structures. The model provides a plausible and quantitative explanation for several unexplained phenomena observed at the tissue level in cardiac muscles. Model execution entails Monte-Carlo-based simulations of Markov representations of calcium regulation and actin-myosin interactions. The model is computationally expensive and requires a supercomputer to simulate sub-cellular structures. While useful to understand biophysical questions, such models are obviously impractical to model the billions of cells that comprise a whole human heart. We have also developed a more computational efficient model that approximated the spatial interactions at the protein level without explicit computation. The goal of this work is to bridge from cells to large organ-level anatomical structures with practical run times. We hope that the power of this approximate model to recapitulate complex force responses in cardiac tissue will foster wider use of cardiac models for research and clinical applications. The work is a case study in multiscale biological modeling where the development of a complex, detailed model is required to guide the development of more abstract and computationally efficient representations.

1 Introduction

Many cardiac phenomenon emerge where effects span extreme spatial and time scales. For example, many cardiac drugs operate on the molecular scale with effects on ionic channels, whereas one wants to understand the effects of these drugs on arrythmias and sudden cardiac death (i.e., the effects at organism and whole heart level over a much longer time scale). Hence, the cardiac field requires models that can span large spatial and temporal scales[1]. The talk presented a body of work around several multiscale models of contraction in heart that attempt to represent the system at different levels of abstraction. The first model is a higly detailed representation of the molecular interactions on a single pair of filaments. This level of detail is required because important controversies exist as to the fundamental molecular mechanisms and how to represent these mathematically[2]. Later, the single filaments are combined into a larger structure called a myofibril. The detailed model also guides the development of abstractions that are computationally efficient and based on ordinary differential equations (ODEs).

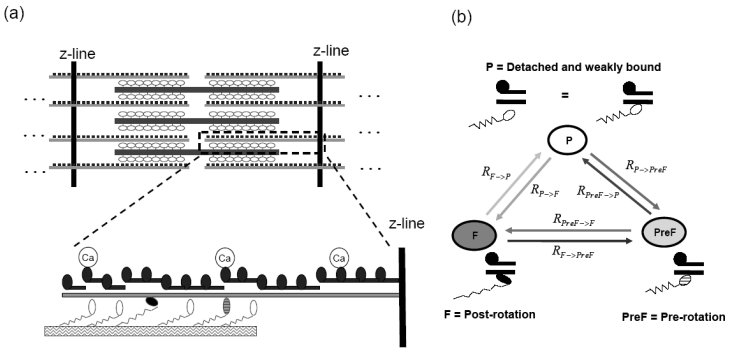


Figure 1. Sarcomere structure in striated muscle. (a) Schematic representation of the repeating sarcomere structure in striated muscle. The sarcomere is defined from z-line to z-line with interdigitated thick and thin filaments that can interact to produce force. (b) State diagram for crossbridge cycling.

2 Mechanistic Model of the Myofilaments

A detailed representation of proteins and protein complexes with spatial interactions is developed to understand fundamental mechanisms in heart contraction[5]. Here the proteins cannot be tractably modeled on first principles based on atomistic approaches such as molecular dynamics. Instead, a more abstract formulation is used. Specifically, the interactions of actin and myosin within one pair of thick and thin filaments in the cardiac sarcomere are represented in Fig. 1a. The sarcomere is the basic repeating of the contractile apparatus in striated muscle. Hence, the model represents a small but repeating structure so that results can be extrapolated to compare with complete muscle responses if one assumes all sarcomeres act equivalently, an assumption that is roughly true for some conditions.

The model includes explicit representations of actin, myosin, and regulatory proteins. For the sake of brevity, we will will focus on the actin and myosin interactions. The thin filament is a two-stranded helix of actin (see Fig. 2a). Myosin has three major structural subunits - the head, neck and tail. The head attaches to a binding site and rotates using the energy from the conversion of adenosine tri-phosphate (ATP) to adenosine di-phosphate (ADP) and inorganic phosphate (P_i), a common energy-liberating reaction in cells. The tail regions of myosin assemble together to form the thick filament (see Fig. 2a). The bound linkages between actin and myosin are commonly termed crossbridges to reflect bridging between thick and thin filaments. Figure 1b shows a Markov model for the crossbridge cycle. State **P** is a detached crossbridge which corresponds to two biochemical states: completely separate and a transient, electrostatic interaction known as weakly bound. State $\mathbf{PreF}$ stands for Pre-Force and corresponds to a more strongly bound state in which the head has not rotated yet. Rotating the head can stretch the extensible neck region to generate force as represented by State $\mathbf{PostF}$ that stands for Post-Force.

The model contains multiple instances of myosin and the corresponding binding sites as adapted from the work of Daniel and colleagues[3]. We have used this approach to form the spatial layout of myosin and actin binding sites with the appropriate compliances be-

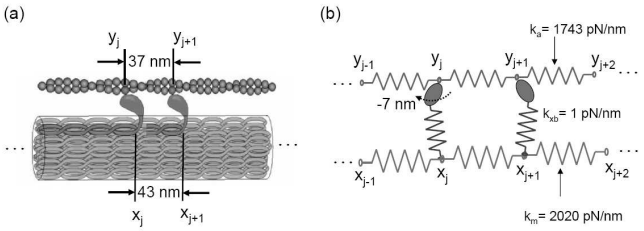


Figure 2. Spatial arrangement of actinmyosin interactions. (a) Myosin heads on the thick filament have an intrinsic spacing of 43 nm, slightly larger than the 37-nm effective spacing of appropriately aligned actin binding sites on the thin filament. (b) Compliances in the thick and thin filaments represented as a system of springs between adjacent binding sites. Attached crossbridges are shown as springs linking the two filaments. Values for actin binding sites (y_i) and myosins (x_i) are computed and modeled as a system of linear springs. Rotation of myosin head changes stretch of crossbridge spring by 7 nm (see text for details).

tween the elements. We assume that the thick filament has myosin heads with appropriate orientations at an intrinsic spacing of 43 nm, as shown in Fig. 2a. In real myosin, the heads extend in a helical fashion such that only a subset will appropriately align to interact with the single thin filament assumed in this model. Note, however, that the 43-nm intrinsic spacing of myosin is slightly larger than the 37-nm spacing of appropriate binding sites on the thin filament. Similarly to the case with myosin, the helical nature of the actin will restrict binding to a subset of actin monomers that appropriately face the thick filament in the two-filament model presented here.

The transition rates between the crossbridge states in Fig. 1b are determined by energy profiles as defined elsewhere[5]. Briefly, the energy profile represent a coupled chemical and mechanical system so that the energy of the hydrolysis of ATP is assumed to permit attachment and rotation of the mysosin head. The model correspond to molecular-level events so both forward and reverse rates are assumed. The rates depend on the relative positions of the actin binding sites (y_i) and myosin (x_i) for a given pairing. The attached states ($\mathbf{PreF}$ and $\mathbf{PostF}$) have a parabolic energy profile that corresponds to a spring element. The rates may also depend on metabolite concentration (ATP, ADP, and P_i), although these are not varied in the current version of the model. At each time step in the Monte Carlo simulation, the state of each interacting pair of actin binding site and myosin are updated. The spatial positions of the sites are calculated by solving the linear system of springs for the whole ensemble (Fig. 2b shows a small subset with only two pairs). Note that the rotation of the head (from State $\mathbf{PreF}$ to $\mathbf{PostF}$) is assumed stretch of crossbridge spring by 7 nm. Likewise, reverse rotation (from State $\mathbf{PostF}$ to $\mathbf{PreF}$) decreases the stretch. Different intrinic spacing of actin binding sites and myosin will produce a distribution of relative positions and transition rates. Moreover, the net motions of the thick and thin filaments that occur during contraction will continually change the relative spacings actin sites and myosin.

The interaction of actin and myosin is controlled by regulatory proteins that binds calcium (Ca) ions. The regulatory proteins (specifically known as troponin and tropomyosin) reside in the two grooves of the two-stranded actin helix and serve to allosterically block interactions between actin and myosin. The term ”allosteric” refers to a change in shape and activity of a protein that results from molecular binding with a regulatory substance.

In the lower half Fig. 1a, we represent a single one-dimension lattice of regulatory units that reside on one of the two grooves. The allosteric shift is schematically illustrated by showing some units raised above the thin filament. The important feature of troponin is that it can bind one Ca ion at a single binding site that controls the allosteric switching behavior. The troponin and tropomyosin molecules overlap in an end-to-end fashion and are thought to communicate with their neighbors via this physical communication, a phenomenon called cooperativity. Hence, the regulatory proteins tend to switch on and off in unison with their neighbors. The communication is critical to produce a steep Ca sensitivity in that a small change in Ca level produces a large change in activation level. While space does not permit more description, the details can be found elsewhere[2,5].

3 Expanding to Terascale Models

One method of generating multiscale models is to use large computers to solve many replicates of the fundamental structure. Along these lines, the basic model (one thin filament and half of thick filament) just described will be expanded to produce a model of the myofibril (a common experimental preparation that can be disected from cells). Figure 3 shows a possible mapping of 32 sarcomeres onto one rack of a Blue Gene/L. The simulation of two thick and eight thin filaments is executed on a dual-core processor. Each thick filament is surrounded by six thin filaments, while each thin filament is surrounded by three thick filaments. This gives a filament ratio of two to one, which must be increased to a ratio of four thin filaments to one thick filament in order to account for the left and right sides of a sarcomere. Note that the thick filaments are double-ended and require roughly twice the computation of the half thick filament in the preliminary model. We anticipate a 64-thick and 256-thin filament to represent a full sarcomere at the level of a computer node card. A myofibril model can then comprise 32 full sarcomeres modeled at the level of a full rack of Blue Gene/L. The mapping is approximate because the final implementation may require some redistribution of the model among computation units at the level of processors or node cards.

One challenge to implement large models on large-scale parallel computers is to efficiently use the hardware in a distributed fashion. One method to is to use existing libraries that allow for distributed processing of sparse matrices. We are using PETSc (Portable, Extensible Toolkit for Scientific Computation)[6]. To solve for the locations of the binding sites, the system $\mathbf{A} \bullet \mathbf{X} = \mathbf{k}$ must be solved at each time step where there is a change in state of the crossbridge attachment. PETSc provides a convenient method to construct the sparse matrices and solve the system in a parallel fashion.

4 Approximate Model Based on Ordinary Diffential Equations

We have also developed a more computational efficient model that approximated the spatial interactions at the protein level without explicit computation of the spatial interactions[4]. The goal of this work is to bridge from cells to large organ-level anatomical structures with practical run times. Conceptionally, this approximate modeling can be compared to coarse-grain methods in protein folding that are developed be much more efficient than atomistic molecular dynamics approaches. However, one must understand and carefully

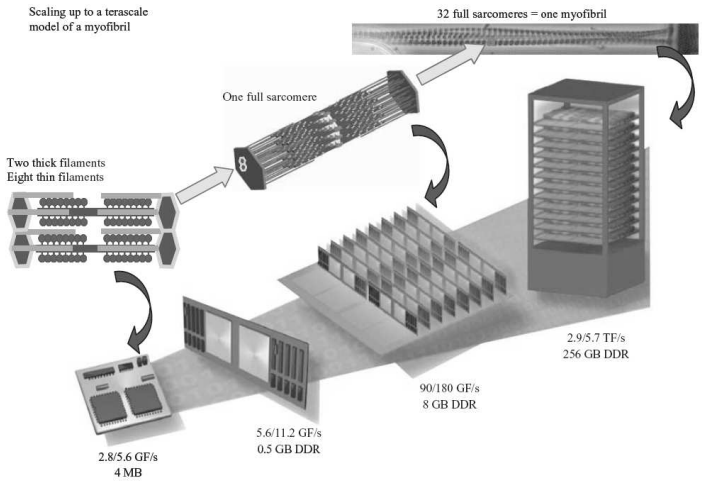


Figure 3. Possible mapping of a 32-sarcomere myofibril model onto one rack of a Blue Gene/L supercomputer. Simulation of two thick and eight thin filaments is executed on a dual-core processor; then, 64 thick and 256 thin filaments can represent a full sarcomere at the level of a node card. The mapping is approximate because a final implementation may require some redistribution of the model among computation units at the level of processors or node cards, in order to balance computational loads, given communication constraints. (DDR: double-data-rate synchronous dynamic random access memory; GB: gigabytes; GF: gigaflops; TF: teraflops.)

weigh the benefits of these approaches as the coarse graining typically trades accuracy and generality for a faster approximate method that works under limited conditions.

The method of transferring a model that is spatially explicit to one that is essentially a non-spatial point model involves representing the mass action of populations rather than indivividual elements. Such an approach can be done with high accuracy when the individual events are indepedent. For example, populations of independent membrane-bound channels can be represented in high accuaracy as probabilities of state cccupancies. In this way, Markov-state diagrams (similar to that in Fig. 1b) can translated diretly into systems of ODEs. However, the case is not so simple for muscle in which many of the important behavior are thought to emerges from the interaction between neighboring entities[2]. The exact method the tranlating to a system of ODEs is given elsewhere[4] ; however, a brief description is provided here. The ODE-based model represents the populations of crossbridges in the states shown in Fig. 1b. In addition, the mean strains of the attached crossbridge states (**PreF** and **PostF**) are computed using phenomenological formulations. The force is computed at mean occupancies multiplied by the mean stain of the attached states with the assumption of an ideal spring constant. This is an approximation as the real system will have attachment rates and strains corresponding to individual crossbridges. Hence, the calculation of the mean values (a construction often termed ”mean field”) is an approximation that suffices under restricted conditions.

We hope that the power of this approximate model to recapitulate complex force responses in cardiac tissue will foster wider use of cardiac models for research and clinical applications. Indeed, the talk presented some preliminary results in a coupled electomeachanical model by Viatcheslav Gurev and Natalia Trayanova at The Johns Hopkins

University. In the whole heart, the ODE-based model provides the mapping between length, shortening velocity, and force during contraction. While preliminary, the results illustrates the work required to developed multiscale models can pay off in improved accuracy at larger scales. In this study, the diffences in timing between contraction at inside (endocardium) and outside (epicardium) of the heart could be recapitulated. Similar results have not been reported previously with phenomenological modeling approaches.

5 Concluding Remarks

The work is a case study in multiscale biological modeling where the development of a complex, detailed model was required to guide the later development of a more abstract and computationally efficient representation. The detailed model seeks to represent spatial interactions between protein complexes that are thought to produce characteristic cardiac muscle responses at larger scales. We have also developed a more computationally efficient model that approximates the spatial interactions at the protein level without explicit computation. The goal of this work is to bridge from cells to large organ-level anatomical structures with practical run times. We hope that the power of this approximate model to recapitulate complex force responses in cardiac tissue will foster wider use of cardiac models for research and clinical applications.

Acknowledgments

Several researchers made substatial contributions to the work described. Jagir Hussan at the University of Auckland and Pieter de Tombe at The University of Illinois Chicago are the coworkers for the mechanistic and ODE-based models of the myofilament. Lei Jin contributed by providing the illustrations of the protein complexes in Fig. 2a. Viatcheslav Gurev and Natalia Trayanova at The Johns Hopkins University contributed results for an electro-mechanical model of whole heart.

References

1. P.J. Hunter and T.K. Borg *Integration from proteins to organs: the Physiome Project*, Nat Rev Mol Cell Biol. **4(3)**, 237-43, 2003.
2. J. J. Rice and P. P. de Tombe *Approaches to Modeling Crossbridges and Calcium-Dependent Activation in Cardiac Muscle*, Prog. Biophys. Mol. Biol. **85, No. 2-3**, 179-195, 2004.
3. T. L. Daniel, A. C. Trimble, and P. B. Chase *Compliant Realignment of Binding Sites in Muscle: Transient Behavior and Mechanical Tuning*, Biophys. J. **74, No. 4**, 1611-1621, 1998.
4. J. J. Rice, F. Wang, D. M. Bers, and P. P. de Tombe *Approximate model of cooperative activation and crossbridge cycling in cardiac muscle using ordinary differential equations*, Biophys. J., 2008, in press.
5. J. Hussan, P. P. de Tombe, and J. J. Rice *A spatially detailed myofilament model as a basis for large-scale biological simulations*, IBM Journal of Research and Development **50(6)**, 2006.
6. `http://www-unix.mcs.anl.gov/petsc/petsc-as/`.

Determination of Protein Structure in Solution Based on $^{13}C^{\alpha}$ Chemical Shifts and NOE Distance Constraints

Jorge A. Vila[1,2] **and Harold A. Scheraga**[1]

[1] Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, U.S.A.
E-mail: has5@cornell.edu

[2] Universidad Nacional de San Luis, Instituto de Matemática Aplicada San Luis, CONICET, Ejército de Los Andes 950, 5700 San Luis, Argentina

A recently introduced physics-based method that exploits distance constraints derived from Nuclear Overhauser Effect and $^{13}C^{\alpha}$ chemical shift data aimed at determining, validating and refining, protein structures at a high level of accuracy, without resorting to other experimental data, is illustrated here by determining the native structures of two proteins, namely a 20-residue all-β and a 76-residue all-α protein. The approach makes use of $^{13}C^{\alpha}$ chemical shifts, computed at the density functional level of theory to derive the backbone and side-chain torsional constraints for *all* backbone and side-chain torsional angles *dynamically*. Consequently, this method is expected to lead to a more precise characterization of the conformational distributions for the backbone, as well as for the side chains of the amino acid residues on both the surface and in the interior of a protein. With the available computational resources, there are three main advantages of this new methodology: (a) it can be used for proteins of *any* class or size; (b) it provides a unified, self-consistent, methodology to determine, validate and refine proteins structures at a *high-quality* level; and (c) it does not use any knowledge-based information and hence, it is a purely *physics-based* method. The anticipated results of our applications indicated that, starting from randomly generated conformations, the final protein models are more accurate than existing NMR-derived models (obtained by using traditional methods) in terms of the agreement between predicted and observed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts as well as some stereochemical quality indicators.

1 Introduction

The backbone and side-chain conformations of a residue are influenced by interactions with the rest of the protein but, once these conformations are established by these interactions, the $^{13}C^{\alpha}$ chemical shift of this residue depends, mainly, on its backbone,[1] and its side-chain[2,3] conformation, with no significant influence of either the amino acid sequence[4] or the position of the given residue in the sequence[4]. These properties, together with the fact that $^{13}C^{\alpha}$ is ubiquitous in proteins, make this nucleus an attractive candidate for computation of theoretical chemical shifts at the quantum chemical level of theory in order to determine, validate and refine protein structures[4,5]. We have been developing methodology to use $^{13}C^{\alpha}$ chemical shifts, in addition to other NMR data, to determine, validate and refine protein structure in solution[5–7].

This methodology[4], validated on 139 conformations of the human protein ubiquitin, enabled us to offer a new criterion for an accurate assessment of the quality of NMR-derived protein conformations and to examine whether X-ray or NMR-solved structures are better representations of the observed $^{13}C^{\alpha}$ chemical shifts in solution. A detailed analysis[4] of the disagreement between observed and DFT-computed $^{13}C^{\alpha}$ chemical shifts in these ubiquitin conformations illustrated the accuracy of the calculations and, more

important, demonstrated that these disagreements reflect the dynamic nature of the protein, rather than inaccuracies of the method. Our methodology has also been used[6] to show that neutral, rather than charged, basic and acidic groups are a better approximation of the observed $^{13}C^{\alpha}$ chemical shifts of a protein in solution.

The goal of this work is to illustrate how this methodology (a) can be used for proteins of *any* class or size; (b) provides a unified, self-consistent, methodology to determine, validate and refine proteins structures at a *high-quality* level; and (c) does not use any knowledge-based information and hence, it is a purely *physics-based* method. To accomplish this goal, the methodology is illustrated here with two applications: first, to determine an accurate set of conformations that simultaneously satisfies the NOE-derived distance constraints and the $^{13}C^{\alpha}$-derived torsional constraints for a 20-residue peptide capable of forming a three-stranded antiparallel β-sheet in aqueous solution[8], i.e., the BS2 peptide with the sequence: TWIQN$_D$PGTKWYQN$_D$PGTKIYT, for which complete sets of both $^{13}C^{\alpha}$ chemical shifts and NOEs were reported[8]. Secondly, as an additional test of the procedure, we chose to determine the tertiary structure of the B. Subtilis Acyl Carrier (SAC) protein[5]. This is a small all-α helical protein with only 76 amino acid residues and no disulfide bonds, for which all the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts and the NOE-derived distances are available from the Biological Magnetic Resonance Data Bank under *accession number* 4989. The NMR structure of the SAC protein has been solved by Xu et al.[9] using traditional methods, and the coordinates of the average-minimized structure were deposited in the Protein Data Bank with the code 1HY8.

2 Materials and Methods

Figure (1) shows a flow chart of the protein structure determination procedure which, essentially, consists of 4 steps, namely:

(1) The Variable-Target-Function (VTF) approach with a simplified soft-sphere potential function[10] is used to generate an ensemble of conformations at random that simultaneously satisfy a set of distance constraints derived from the experimental NOEs and the backbone torsional constraints derived from the $^{13}C^{\alpha}$ conformational shifts, i.e., only for the regular α-helical and β-sheet segment of the molecule. Among all generated VTF conformations, only those possessing a maximum NOE-derived distance violation lower than a certain cut off value, e.g., 1 Å, are selected. If the number of selected conformations is greater than $\sim$10, then a clustering procedure is applied by using the Minimal Spanning Tree (MST) method[11].

(2) The $^{13}C^{\alpha}$ chemical shifts are computed at the DFT level[4–6] for each conformation of the set obtained in step (1). The DFT procedure is applied to each amino acid **X** in the sequence by treating **X** as a terminally-blocked tripeptide with the sequence Ac-G**X**G-NMe in the conformation of each generated peptide structure. Examination of the chemical shifts of each residue of all the clustered conformations considered here enabled us to identify a new *minimal*-rmsd model[4] in which the $^{13}C^{\alpha}$ chemical shift of each residue individually best matched the experimental one, thereby providing a *new* set of ϕ, ψ, and χ torsional angle constraints of *all* the residues of the molecule[5].

(3) Only one conformation among all the selected conformations described in step (1), was selected. This conformation possessed the lowest rmsd between the computed and observed $^{13}C^{\alpha}$ chemical shifts. The selected conformation was used as a starting one in a

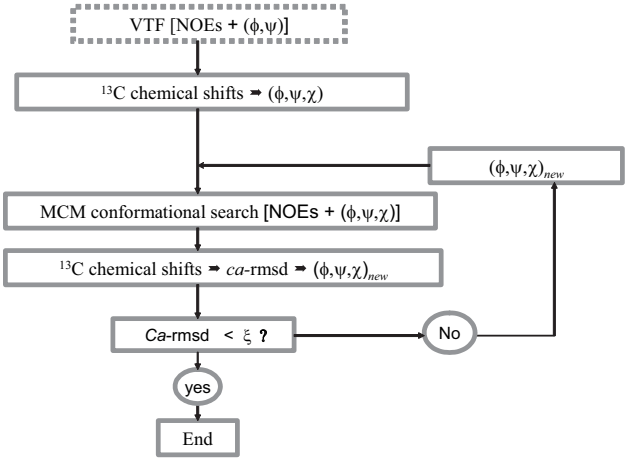


Figure 1. Flow chart illustrating the steps of the computational procedure, as described in the Materials and Methods section. VTF is the acronym for the Variable-Target-Function approach[10]. The variable ξ represents the convergence criterion.

conformational search with the Monte Carlo with Minimization (MCM)[12] method carried out with two types of constraints: the original fixed set of NOEs and the *new* set of torsional angles derived in step (2). This time, instead of using a simplified soft-sphere potential function, we use a complete force-field containing the following terms: (a) the internal potential energy, as described by the ECEPP/3 force field[13]; and (b) additional energy terms aimed at penalizing violations of the distance and torsional angle constraints[14]. Finally, a clustering procedure is carried out to select a small sub-set of the total number of the MCM-derived set of conformations by using the MST method[11] and assuming a specific rmsd cutoff for all heavy atoms.

(4) Steps (2) and (3) are repeated iteratively by using the set of conformations obtained in step (3) and, hence, enabling us to obtain an *updated* set of torsional-angle constraints. At any stage of the procedure, a tolerance range Λ, with $20^o \leq \Lambda \leq 35^o$, for the torsional constraints was adopted. Variation of the torsional angles within a tolerance range Λ is considered acceptable and hence is not subject to energetic penalties. Among all the conformations generated in the final use of step (3), only one conformation is selected, because it is characterized by the lowest rmsd between the computed $^{13}C^{\alpha}$ chemical shifts and the observed ones. Thus, the procedure of step 3, applied to such a conformation, led to a new set of structures. The final number of conformations in this set is determined by the cutoff rmsd value adopted for the clustering procedure in step (3).

Application of this procedure to 20-residue and 76-residue proteins enabled us to determine a Set_β, consisting of 10 conformations[7], and a Set_α, consisting of 9 conformations[5], respectively. Analysis of the quality of these sets of conformations in terms of the *ca*-rmsd[4] and some structural quality indicators are given in Table 1.

Conformation Set[a]		*Ca*-rmsd[b] [ppm]	Maximum Distance Violation[c] [Å]	Structural Quality Indicators[d]
all-β	Santiveri (20)[8]	4.6	2.36	61 ± 11 (39 ± 12) [0.0]
	Set_β (10)[7]	3.5	0.88	62 ± 10 (37 ± 9) [0.0]
all-α	1HY8 (1)[9]	3.9	1.38	95.8 (2.8) [1.4]
	Set_α (9)[5]	2.9	0.63	92.0 ± 1 (4.3 ± 0.9) [1.5]

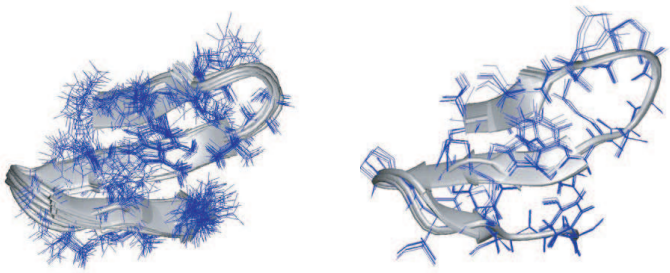
Table 1. Results for the all-α and all-β protein structure determination.

[a] Computed for each set of conformations listed; the number of conformations in each set is indicated in parentheses. Set_β[7] and Set_α[5] refer to the set of conformations determined as explained in the Materials and Method section.
[b] Values computed as explained in Vila et al.[4]
[c] From the *full* set of NOE-derived distances, namely 130 for the 20-residue all-β and 1,050 for the 76-residue all-α proteins, respectively.
[d] Based on PROCHECK[15]. The listed values are the number of residues in the allowed regions of the Ramachandran map; in the generously allowed regions (in parenthesis); and in the disallowed regions (in bracket). All the listed values, except for the protein 1HY8, are averaged over the total number of conformations of each set.

3 Results and Discussion

The results obtained here indicate that an *all* β-sheet (Figure 2a-b) and an *all* α-helical (Figure 3) set of structures can be determined by simply identifying a set of conformations which simultaneously satisfy a set of constraints, namely $^{13}C^{\alpha}$-dynamically-derived torsional angle constraints for *all* amino acid residues in the sequence, and a fixed set of NOE-derived distance constraints. Analysis of the accuracy of these sets, as a measure of the closeness with which the calculations reproduce the structure in solution, in terms of the NOE-derived maximum distance violations, the $^{13}C^{\alpha}$ chemical shifts, and some stereochemical quality factors (see Table 1), indicates that our self-consistent physics-based method is able to produce a more accurate set of conformations than that obtained with the traditional methods.

In summary, these applications illustrate the three main advantages of this new methodology: (a) it can be used for proteins of *any* class or size; (b) it provides a unified, self-consistent, methodology to determine, validate and refine protein structures at a *high-quality* level; and (c) it does not use any knowledge-based information and, hence, it is a purely *physics-based* method.



(a) **(b)**

Figure 2. (**a**) Superposition of 20 NMR-derived conformations (represented by ribbon diagrams) of the BS2 peptide obtained by Santiveri *et al*[8]. Side chains are represented by thin black lines. (**b**) Same as (**a**) for the 10 NMR-derived conformations in this work from Set_β[7].

Figure 3. Ribbon diagram of the superposition of nine models of the Set_α for the SAC protein[5] (in grey color) and the minimized average NMR structure (1HY8) [in black][9].

Acknowledgments

This research was supported by grants from the National Institutes of Health (GM-14312 and GM-24893), and the National Science Foundation (MCB05-41633). Support was also received from the CONICET, FONCyT-ANPCyT (PAV 22642 / 22672), and from the Universidad Nacional de San Luis (P-328501), Argentina. The research was conducted using the resources of a Beowulf-type cluster located at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University; and the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center.

References

1. S. Spera and A. Bax Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C Nuclear Magnetic Resonance chemical shifts. J. Am. Chem Soc., 113: 5490-5492, 1991.
2. R.H Havlin, H. Le, D.D. Laws, A.C. deDios and E. Oldfield. An ab initio quantum chemical investigation of carbon-13 NMR shielding tensors in glycine, alanine,

valine, isoleucine, serine, and threonine: Comparisons between helical and sheet tensors, and effects of χ_1 on shielding. J Am Chem Soc., 119: 11951-11958, 1997.
3. M.E. Villegas, J.A. Vila and H.A. Scheraga. Effects of Side-Chain Orientation on the ^{13}C Chemical Shifts of Antiparallel β-sheet Model Peptides. J Biomol NMR, 37: 137-146, 2007.
4. J.A. Vila, M.E. Villegas, H.A. Baldoni and H.A. Scheraga. Predicting $^{13}C^{\alpha}$ chemical shifts for validation of protein structures. J. Biomol. NMR, 38:221-235, 2007
5. J.A. Vila, D.R. Ripoll and H.A. Scheraga. Use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination. J. Phys. Chem. B, 111:6577-6585, 2007.
6. J.A. Vila and H.A. Scheraga. Factors affecting the use of $^{13}C^{\alpha}$ chemical shifts to determine, refine, and validate protein structures. Proteins 71**,** 641-654, 2008.
7. J.A. Vila, Y.A. Arnautova, and H.A. Scheraga. Use of $^{13}C^{\alpha}$ chemical shifts for accurate determination of β-Sheet structures in solution. Proc. Natl. Acad. Sci. USA**,** 105, 1891-1896, 2008.
8. C.M. Santiveri, J. Santoro, M. Rico and M.A. Jiménez. Factors involved in the stability of isolated beta-sheets: turn sequence, beta-sheet twisting, and hydrophobic surface burial. Prot. Sci., 13:1134-1147, 2004.
9. G.-Y. Xu, A. Tam, L. Lin, J. Hixon, C.C. Fritz and R. Powers. Solution structure of B. Subtilis Acyl carrier protein. Structure, 9:277-287, 2001.
10. M. Vásquez and H.A. Scheraga. Variable-Target-Function and buildup procedures for the calculation of protein conformation – application to bovine pancreatic trypsin-inhibitor using limited simulated Nuclear Magnetic-Resonance data**.** J. Biomol. Struct. Dyn., 5:757-784, 1998.
11. J.B. Kruskal, Jr., On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proc. American. Math. Soc., 7:48-50, 1956.
12. Z. Li and H.A. Scheraga. Monte-Carlo-Minimization approach to the multiple-minima problem in protein folding. Proc. Natl. Acad. Sci. USA, 84:66116615, 1987.
13. G. Némethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H.A. Scheraga. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J. Phys. Chem., 96:6472-6484, 1992.
14. D.R. Ripoll and F. Ni. Refinement of the thrombin-bound structure of a hirudin peptide by a restrained Electrostatically Driven Monte-Carlo Method. Biopolymers, 32:359-365, 1992.
15. R.A. Laskowski, M.W. MacArthur, D.S. Moss and J. Thornton. PROCHECK - a program to check the stereochemical quality of protein structures. J. Appl. Cryst.,26:283-291, 1993.

Computer Simulation of Biomolecular Systems: Where Do We Stand?

Wilfred F. van Gunsteren and Daan P. Geerke

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH,
8093 Zurich, Switzerland
E-mail: wfvgn@igc.phys.chem.ethz.ch

The four major aspects that determine the quality of the ensemble of molecular conformations as obtained from biomolecular simulation are reviewed and illustrated with examples.

1 Introduction

Over the past three decades, simulation of the dynamics of biomolecular systems at the atomic level has developed from short-time simulations of simple molecular models [1, 2] to orders of magnitude larger simulations based on detailed and much more accurate molecular models [3]. The improved accuracy has turned molecular dynamics (MD) simulation into a standard method for an atomic interpretation of experimental data on biomolecular systems [4]. Yet, much progress is still to be made in order to use MD simulation to accurately predict various properties of biomolecular systems [5].

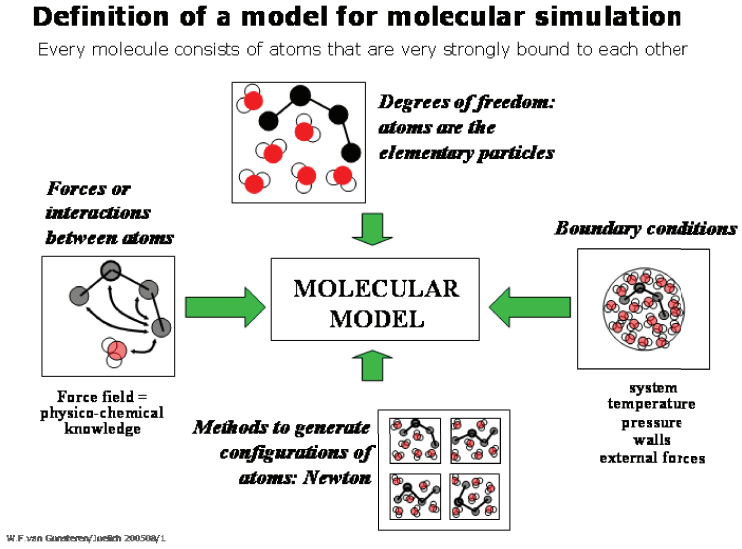


Figure 1.

Figure 1 illustrates the four choices to be made when defining a molecular model for molecular simulation: (1) which degrees of freedom are to be explicitly simulated; (2) how are the forces governing the motion along these degrees of freedom calculated in a necessarily approximative manner; (3) how is the motion of the system propagated in time such that the relevant configurational space is widely and sufficiently sampled; (4) how are the spatial and thermodynamic boundary conditions imposed upon the motion of the system. In this short paper we illustrate the state of the art with respect to these four aspects of modelling using examples from our own work.

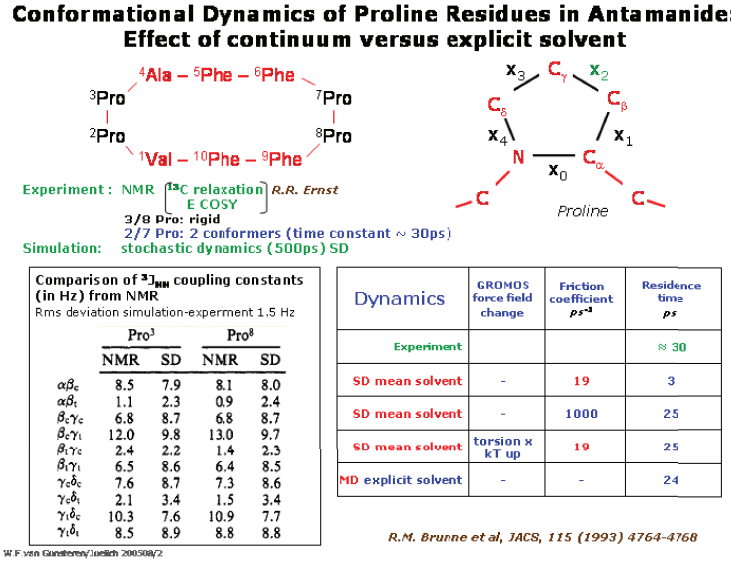


Figure 2.

2 Choice of Degrees of Freedom

In Figure 2, the importance of explicitly including solvent degrees of freedom is illustrated. It compares the puckering residence time of the χ_2 torsional angle of residues ^{2}Pro and ^{7}Pro in the cyclic polypeptide antamanide as obtained from experiment [6] with that as obtained from simulations using either a mean implicit solvent [7] or explicit water solvent molecules [8]. Using stochastic dynamics (SD) simulation with a friction coefficient of 19 ps^{-1}, typical for water at room temperature and pressure, the puckering rate is ten times too high compared to experiment. Using a mean solvation model it could only be reduced by either using an artificially high friction coefficient of 1000 ps^{-1} or by artificially increasing the torsional barrier by about 2.5 kJ mol^{-1}. These unjustified changes in the molecular model can, however, be avoided by explicitly simulating the water degrees of freedom, which results in the correct puckering rate (see Figure 2).

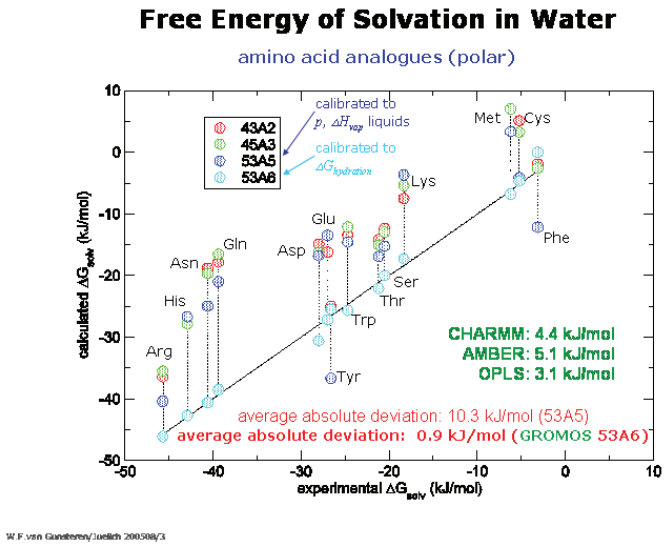


Figure 3.

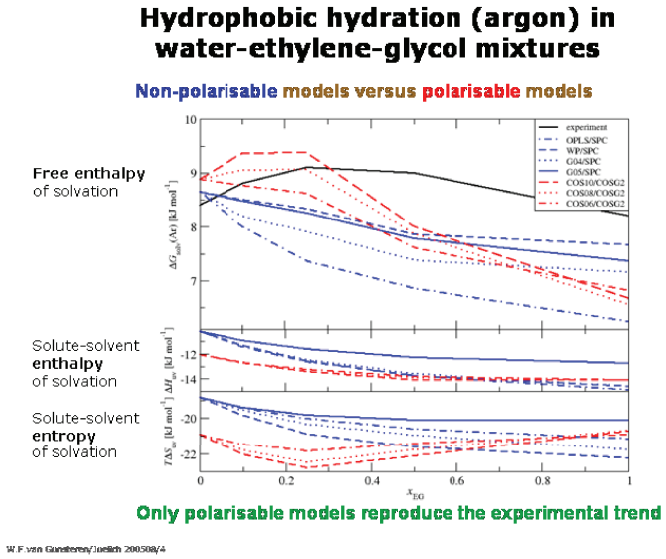


Figure 4.

3 Choice of Interatomic Interactions

In Figure 3, it is illustrated that the free energy of aqueous solvation of the side chains of polar amino acid residues is insufficiently negative for some widely used biomolecu-

lar force fields [9], which would lead to incorrect partitioning between polar and apolar solvents and would over-stabilize folded protein structure. Therefore, more recent force fields, such as the GROMOS 53A6 one [10], lead to a better description of the folding equilibrium [11].

Figure 4 illustrates that inclusion of atomic polarisability will be essential to reach an improved level of accuracy, i.e. beyond 1 kJ mol $^{-1}$. The experimentally observed non-linear behaviour of the free enthalpy of solvation of argon in water-ethylene glycol mixtures as function of the ethylene glycol mole fraction is only reproduced in MD simulations using polarisable molecular models (COS models), whereas a more or less linear behaviour is obtained when using non-polarisable models [12]. Calculation of the solute-solvent entropy of solvation shows that the non-linearity is an entropic effect, which cannot be modelled using a mean or continuum solvent model.

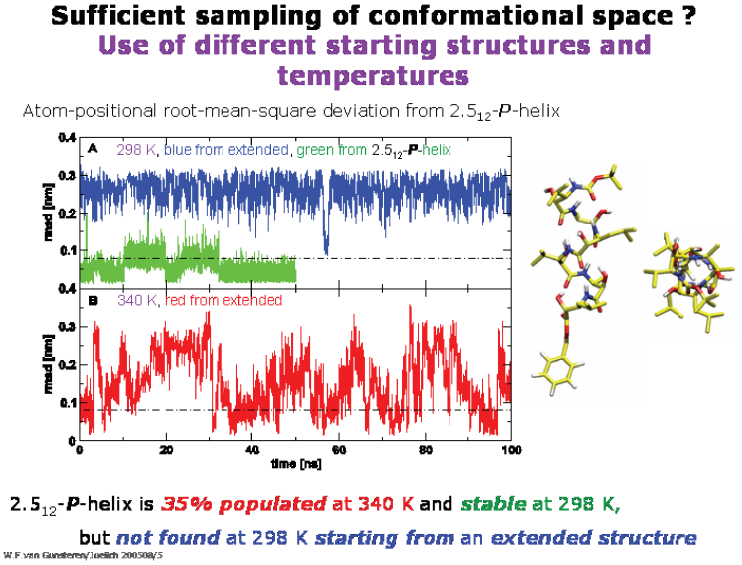


Figure 5.

4 Sufficient Searching and Sampling of Conformational Space

In Figure 5 it is illustrated that even 100 ns of MD simulation of an eight-residue β-peptide in methanol at room temperature may be insufficient to find the most stable 2.5$_{12}$-P-helical fold [13]. Starting the MD simulation at 298 K from this helical structure shows a low root-mean-square deviation (rmsd) from this fold, whereas starting at 298 K from an extended conformation the helical structure is not populated. At 340 K the sampling of conformational space is much widened: a large number of (un)folding events is observed and the helical conformation is present 35% of the time. This example illustrates the need

for search and sampling enhancement techniques, of which Figure 6 classifies the most important ones [14].

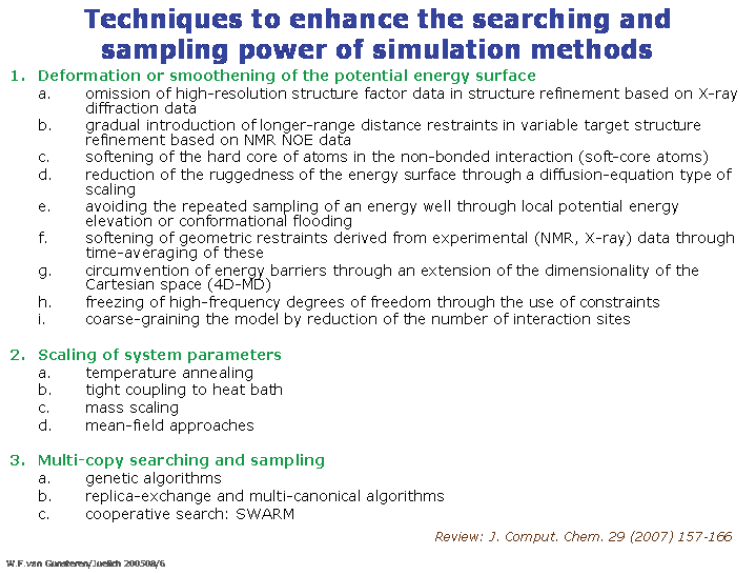


Figure 6.

5 Choice of the Appropriate Thermodynamic State Point

Figure 7 illustrates the effect of pH on the folding equilibrium of a seven residue β-peptide in methanol [15]. The only protonisable groups of this peptide with aliphatic side chains are the amino- and carboxy-termini. The backbone atom-positional root-mean-square deviation of the MD trajectory structures from the most stable (both computationally and experimentally) 3_{14}-helical fold shows that only if the protonation state corresponds to the experimental pH, the helical fold is the most populated one.

6 Conclusion

We have briefly illustrated that the following factors are essential to obtain a high quality ensemble of molecular conformations in a molecular simulation.

1. Inclusion of the relevant degrees of freedom: solvent and co-solvents.
2. Use of a thermodynamically calibrated force field, with a solvent model that is compatible with the solute one, and possibly inclusion of polarisability.
3. Sufficient and Boltzmann-weighted sampling of conformational space.

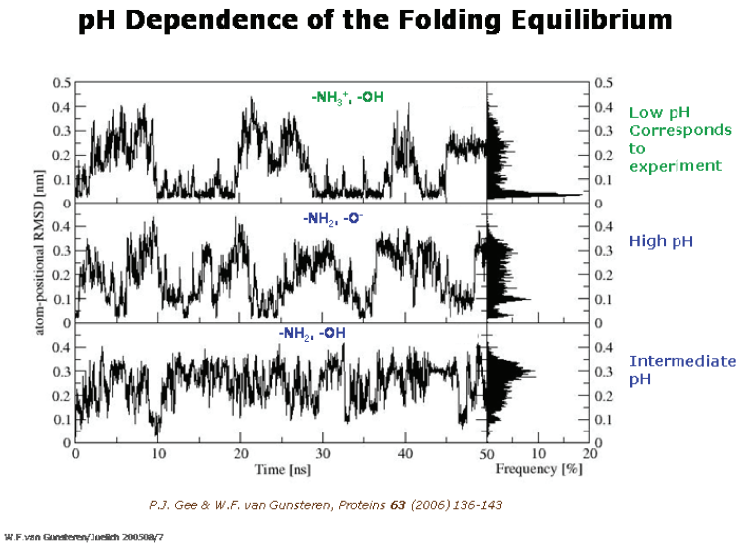


Figure 7.

4. Use of the appropriate (experimental) thermodynamic state point and spatial boundary conditions: temperature, pressure, pH, ionic strength, co-solvents, etc.

Acknowledgments

Financial support from the National Centre of Competence in Research (NCCR) in Structural Biology and grant number 200021-109227 of the Swiss National Science Foundation is gratefully acknowledged.

References

1. A. Warshel, M. Levitt, *Theoretical Studies of Enzymatic Reactions: Dielectric Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme*, J. Mol. Biol. **103**, 227–249, 1976.
2. J. A. McCammon, B. R. Gelin, M. Karplus, *Dynamics of Folded Proteins*, Nature **267**, 585–590, 1977.
3. W. F. van Gunsteren, J. Dolenc, *Biomolecular Simulation: Historical Picture and Future Perspectives*, Biochem. Soc. Trans. **36**, 11–15, 2008.
4. W. F. van Gunsteren, J. Dolenc, A. E. Mark, *Molecular Simulation as an Aid to Experimentalists*, Curr. Opin. Struct. Biology **18**, 149–153, 2008.
5. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glaettli, P. H. Huenenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, H. B. Yu, *Biomolec-*

ular Modelling: Goals, Problems, Perspectives, Angew. Chem. **118**, 4168–4198, 2006; Angew. Chem. Int. Ed. **45**, 4064–4092, 2006.
6. Z. L. Madi, C. Griesinger, R. R. Ernst, *Conformational Dynamics of Proline Residues in Antamanide. J Coupling Analysis of Strongly Coupled Spin Systems Based on E.COSY Spectra*, J. Am. Chem. Soc. **112**, 2908–2914, 1990.
7. R. M. Brunne, W. F. van Gunsteren, R. Brueschweiler, R. R. Ernst, *Molecular Dynamics Simulation of the Proline Conformational Equilibrium and Dynamics in Antamanide Using the GROMOS Force Field*, J. Am. Chem. Soc. **115**, 4764–4768, 1993.
8. J. W. Peng, C. A. Schiffer, P. Xu, W. F. van Gunsteren, R. R. Ernst, *Investigations of Peptide Hydration Using NMR and Molecular Dynamics Simulations: A study of Effects of Water on the Conformation and Dynamics of Antamanide*, J. Biomol. NMR **8**, 453–476, 1996.
9. M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins*, J. Chem. Phys. **119**, 5740–5761, 2003.
10. C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. **25**, 1656-1676, 2004.
11. C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Validation of the 53A6 GROMOS force field*, Eur. Biophys J. **34**, 273–284, 2005.
12. D. P. Geerke, W. F. van Gunsteren, *The performance of non-polarizable and polarizable force-field parameter sets for ethylene glycol in molecular dynamics simulation of the pure liquid and its aqueous mixtures*, Mol. Phys. **105**, 1861–1881, 2007.
13. A. Glaettli, W. F. van Gunsteren, *Are NMR-derived model structures for peptides representative for the ensemble of structures adopted in solution? Probing the fourth helical secondary structure of β-peptides by molecular dynamics simulation*, Angew. Chem. **116**, 6472-6476, 2004; Angew. Chem. Int. Ed. Engl. **43**, 6312-6316, 2004.
14. M. Christen, W. F. van Gunsteren, *On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review*, J. Comput. Chem. **29**, 157–166, 2007.
15. P. J. Gee, W. F. van Gunsteren, *Terminal-group effects on the folding behaviour of selected beta-peptides*, Proteins **63**, 136–143, 2006.

Towards Accurate Predictions of Binding Affinities: The Simulated Scaling Based Free Energy Method

Lianqing Zheng, Hongzhi Li, and Wei Yang

Department of Chemistry and Biochemistry and Institute of Molecular Biophysics,
Florida State University, Tallahassee, FL 32306, USA
E-mail: yang@sb.fsu.edu

1 Introduction

Among various available computational methods, free energy simulation technique [1-7], based on molecular dynamics or Monte Carlos simulations, is unique; this is, if simulation time permits and energy potential is reasonably accurate, it can ultimately lead to quantitative predications of free energy values corresponding to the processes of interest, in particular the binding processes involving drugs and other biologically relevant agents because this method is built upon solid statistical mechanical theories. Unfortunately, with the present computing power and in particular the state-of-the-art free energy simulation algorithms, reaching adequate simulation time for nice free energy convergence, is still quite challenging.

Facing such challenge and also great opportunity, we have been motivated to consider solving these problems. One of our recent developments, the simulated scaling (SS) based method [8], shows intriguing efficiency and robustness. The simulated scaling based free energy simulation method originated from stepwise generalizations of the simulated tempering method. Specifically, at the first step, the temperature space random walk in simulated tempering was generalized to be the potential scaling parameter space random walk based on the modified potential: $U = \lambda U_s + U_e$, where the original energy potential U_0 is decomposed to U_s and U_e, and the scaled energy potential U_s represents the energy terms determining the local conformations of a region of interest. Thereby, the developed algorithm, named by us as the simulated scaling method, allows local sampling to be enhanced in the conformational region described by U_s. In order to realize simultaneous improvement in phase space overlap sampling for free energy simulations, we further generalized the SS method and made it coupled with the dual-topology alchemical free energy simulation setup. Via this generalization, both phase space overlap sampling and conformational sampling problems can be synergistically dealt with. As discussed in our early work, the SS method can also be employed with the single-topology setup, in which A and B share the same set of coordinates; in this setup, conformational sampling problem cannot be ensured as robustly as in the dual-topology setup.

In the following sections, we will describe the SS based free energy method and its extension in the simulations based on quantum mechanical potentials.

2 Simulated Scaling (SS) Method for Localized Enhanced Sampling

For a system with the potential $U_0 = U_s + U_e$, where U_s represents the energy terms determining local conformations in an interested region and U_e represents the rest of en-

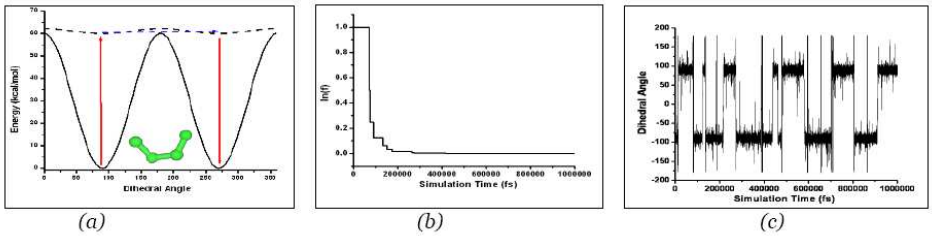
vironmental energy terms. Usually, because of the existence of large energy barriers, converged traveling among various energy minima via regular canonical sampling is time-consuming, sometimes even impossible within currently computer-reachable simulation timescales. To overcome this problem, we can build an expended ensemble with the scaled potential $U = \lambda_m U_s + U_e$, where the dimensionality of the system is extended to 3N+1 (N is the number of real particles) with an additional one-dimension dynamic species λ_m. The canonical traveling in the λ_m space can be realized via hybrid Monte Carlo method [9]. Based on the constructed scaled potential, an acceptance probability for a move from λ_0 to λ_1 can be set as

$$\begin{aligned} p^0_{acpt}[(\lambda_0 \rightarrow \lambda_1)] &= \min\{1, \exp[-\beta[\lambda_1 U_s + U_e) - (\lambda_0 U_s + U_e)]]\} \\ &= \min\{1, \exp[-\beta(\lambda_1 - \lambda_0)U_s]\} = \min\left\{1, \exp\left[- \beta\Delta\lambda\frac{\partial U}{\partial \lambda}\right]\right\}, \end{aligned} \tag{1}$$

in which energy derivative $\frac{\partial U}{\partial \lambda}$ is equal to U_s and scaling parameter change $\Delta\lambda$ is equal to $(\lambda_1 - \lambda_0)$. In this way, the moves in the λ_m space will allow possibly efficient barrier crossing, detoured through the path with λ_m decrease from 1 (effective lowering the energy barriers), barrier crossing (with lower energy barriers), and λ_m increase (returning to 1). However, the probability distribution $\rho(\lambda_m)$ in the λ_m space is determined by the λ_m-dependent free energy profile (potential of mean force), roughness of which very possibly hinders an efficient λ_m move. In order to make a λ_m random walk possible, as discussed in last section, a weight function $a(\lambda_m)$ can be introduced to flatten the λ_m distribution by the application of a biased acceptance probability

$$\begin{aligned} p^{biased}_{acpt}[(\lambda_0 \rightarrow \lambda_1)] &= \min\left\{1, \exp\left(- \beta\Delta\lambda\frac{\partial U}{\partial \lambda}\right)\frac{\exp[a(\lambda_0)]}{\exp[a(\lambda_1)]}\right\} \\ &= \min\left\{1, \exp\left(- \beta\Delta\lambda\frac{\partial U}{\partial \lambda}\right)\frac{f(\lambda_0)}{f(\lambda_1)}\right\}, \end{aligned} \tag{2}$$

in which $\exp[a(\lambda_m)]$ is defined as biasing probability function $f(\lambda_m)$. Thereby, the weight function $a(\lambda_m)$ can be recursively updated with the modification of $f(\lambda_m)$. Specifically, in order to efficiently flatten the λ_m histogram, the updating scheme in the Wang-Landau algorithm [10] is adopted here. When each time a λ_m state is visited after a Monte Carlo acceptance judgment, we update the corresponding biasing probability function $f(\lambda_m)$ using a modification factor $f > 1$, i.e., $f(\lambda_m) \rightarrow f(\lambda_m)/f$. The initial modification factor f_0 can be set as a large value in order for the system to quickly visit all the λ_m states, defined in a certain range $[\lambda_{\min}, \lambda_{\max}]$. This large modification factor f_0 is kept till λ_m random walk results in a "flat" accumulated histogram $H(\lambda_m)$. It is noted that the "flatness" judgment is based on a criteria of whether all the accumulated histograms $H(\lambda_m)$ are not less than a large percentage (80% is often used in the Wang-Landau multi-canonical algorithm) of the average histogram $\langle H(\lambda_m)\rangle$. Then before the next round of λ_m random walk, this modification factor f is reduced to a finer one, updated by a monotonically decreasing function (here $f_{i+1} = \sqrt{f_i}$ is used) and $H(\lambda_m)$ is reset to zero. Following this procedure, a number of cycles will be run continuously till the modification factor is extremely small. In the present development, we can also take a variation by switching



(a) (b) (c)

Figure 1. (a) Torsional potential for the "butane-like" molecule used in model system 1. Black curve represents the unscaled potential for this "butane-like" molecule; dash line represents a scaled potential. (b) Time evolution of during the simulation. (c) Dihedral angle change with the progress of simulation.

this modification factor f to be 1 when it is smaller than a pre-set cutoff value in order to obtain meaningful non-history dependent ensemble. As we will illustrate, the present simulated scaling algorithm shows superior updating efficiency in obtaining converged weight function $a(\lambda_m)$.

Illustrative Example: A "Butane-Like" Molecule. It only has bond, angle and dihedral terms in its potential. Bond and angle terms of this molecule are treated with the CHARMM "Butane" parameters as U_e term. Its dihedral term is set as a double-well potential: $30(1 + \cos(2x))$ kcal/mol, which has a high energy barrier (60 kcal/mol) and so almost exclusively only a λ_m tunneling mechanism (Figure 1a) can enables jumps between the two energy wells at 300 K. Here, the dihedral potential acts as U_s term, which is scaled by λ_m. In the simulated scaling method, facilitated by a λ_m random walk, such barrier crossing is guaranteed, as shown in Figure 1c, through a tunneling mechanism (Figure 1a); this tunneling mechanism [11] is realized by at least three basic steps of moves in the expended coordinate system: a λ_m decrease (solid arrow), physical energy barrier crossing in low λ_m potential (dash arrow), and a λ_m return (solid arrow).

3 Free Energy Simulations with Synergistic Localized Enhanced Sampling Treatments

In the improvement of the free energy simulation techniques, conformational sampling and free energy convergence (even independent of conformational sampling) are two major issues, although these two issues have been mainly discussed as independent topics. There is lack of discussion on how to efficiently deal with these two problems simultaneously, which is required to quickly and reliably obtain the free energy differences (i.e., rather than the precise pseudo-converged values corresponding to certain trapped local conformations).

Usually, an enhanced sampling method is designed to efficiently obtain the ensemble information on certain target physical states. Several accessory states (such as the ones with varied temperatures or Hamiltonians) can be introduced to efficiently propose the structures (to effectively cross the energy barriers) for the target state sampling. Conventionally, the measure of the "conformational sampling" efficiency is purely based on the results from the obtained target state ensembles. However, the accessory states can be use-

ful intermediate states to provide decent phase space overlaps to bridge the two end target physical states for the free energy evaluations. Therefore, we realize that in contrast to the "conformational sampling" problem, the sampling design for the free energy simulation also requires careful consideration of the accessory state ensembles. For free energy simulations, the sampling strategy should be designed with this special concern in mind, because various "conformational sampling" methods may have different level of efficiency in providing an appropriate set of the accessory state ensembles for the evaluation of the free energy difference.

Based on the above thought, an approach is proposed towards two goals: robustly enhanced canonical sampling, for which the accessory state ensembles are designed, and simultaneous quick free energy convergence, facilitated by the decent phase space overlaps provided by these accessory ensembles. Specifically, we propose a dual-topology alchemical simulated scaling (DTA-SS) method, here λ_m plays double roles, viz. 1) as the label of the intermediate state to improve the phase overlap along the alchemical direction; and 2) synergistically as an effective temperature label to enhance the SS sampling efficiency.

Theoretical Design of the Method. The scaled energy function can be rewritten in the dual-topology hybrid potential form, which is usually utilized in free energy simulations, as shown below

$$U = (1 - \lambda_m)U_s^A(\vec{x}) + \lambda_m U_s^B(\vec{x'}) + U_e, \tag{3}$$

where $U_s^A(\vec{x})$ and $U_s^B(\vec{x'})$ represent the unique portions of the energy terms for the two end point chemical species A and B. It should be noted that Equation 3 can also be expressed in a nonlinear form, such as in the form of the soft-core potentials . Therefore, Equation 3 can be generalized as

$$U = f(\vec{x}, \vec{x'}, \lambda) + U_e, \tag{4}$$

in which we have the constraints of $f(\vec{x}, \vec{x'}, 0) = U_s^A(\vec{x})$ and $f(\vec{x}, \vec{x'}, 1) = U_s^B(\vec{x'})$ to recover the chemical end states; the scaled portions have the independent coordinates $\vec{x}$ and $\vec{x'}$, the corresponding potentials of which are scaled in the opposite directions. Here, for simplicity, the discussions will be based on the linear one, Equation 3. As mentioned above, λ and $(1 - \lambda)$ become the labels of the local effective temperatures concomitant with their roles as the potential scaling parameters in the original alchemical free energy simulation design. In this case, the SS algorithm can still be applied, except that here, the energy derivative $\frac{\partial U}{\partial \lambda}$ is equal to $U_s^B(\vec{x'}) - U_s^A(\vec{x})$, when Equation 2 is applied. Consequently, when the λ_m histogram is flattened, free energy difference between any two λ_m states can be naturally obtained according to the following formula

$$\Delta A(\lambda_0 \rightarrow \lambda_1) = -RT[a(\lambda_1) - a(\lambda_0)] = -RT \ln \left(\frac{f(\lambda_1)}{f(\lambda_0)} \right), \tag{5}$$

where $a(\lambda_m)$ and $f(\lambda_m)$ respectively represents the weight function and biasing probability function values. And the time evolution of this computed free energy is expected to behave like in Figure 2b.

Technically, it is very difficult to reach the absolute flatness for the λ_m histogram in order to apply Equation 5, because with the histogram flattened, the modification factor f becomes smaller and its capability to flatten the λ_m histogram is correspondingly reduced. A revised procedure can be used by turning the modification factor f to 1, after it is smaller

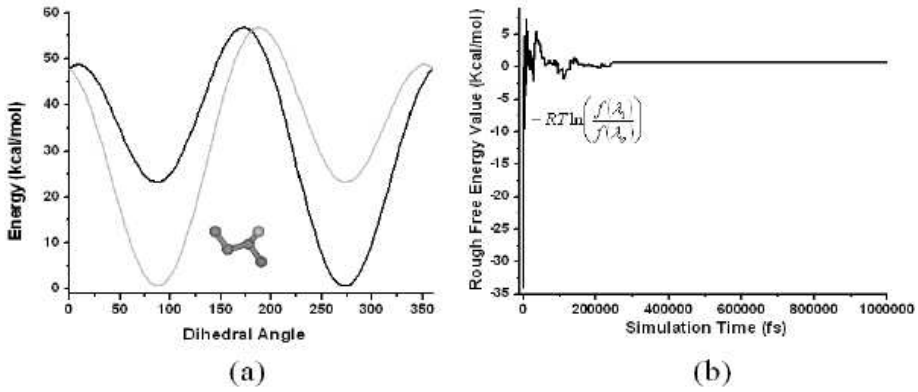


Figure 2. (a) Model potentials set to compute the free energy difference between two "butane-like" molecules. (b) Time evolution of rough free energy, estimated based on Equation 5.

than a pre-set cutoff value. When the modification factor f is 1, the biasing probability function $f(\lambda_m)$ will be constant thereafter and all the ensemble property calculations based on the cumulated histogram $H(\lambda_m)$ are statistically meaningful. Therefore, theoretically exact alchemical free energy value can be estimated using the following equation:

$$\Delta A(\lambda_0 \rightarrow \lambda_1) = -RT \ln \left(\frac{f(\lambda_1)}{f(\lambda_0)} \frac{H(\lambda_1)}{H(\lambda_0)} \right), \tag{6}$$

where $f(\lambda_m)$ is the biasing probability function value for λ_m, when the modification factor f is turned to 1; and $H(\lambda_m)$ is the corresponding accumulated histogram value, counted after the modification factor f is turned to 1.

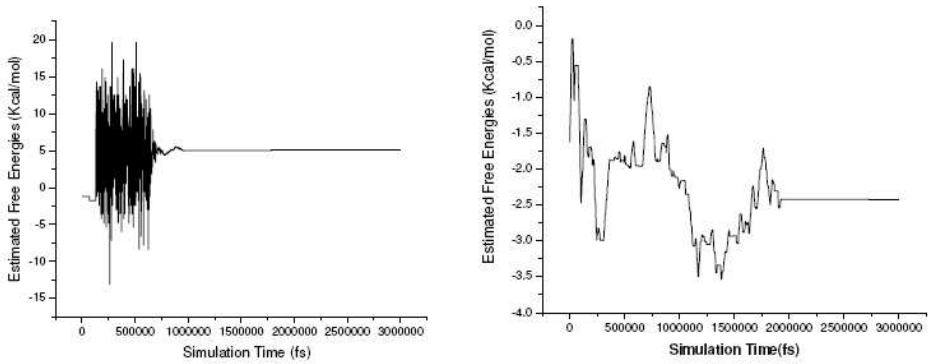
Illustrative Example. For the purpose of demonstrating a simultaneous sampling enhancement accompanying free energy simulation in this method, a model system is set to compute the free energy difference between two end point chemicals with the dihedral potentials of $U = 12\{1 + \cos[\phi \pm 110^0]\} + 20\{1 + \cos(2\phi)\}$. As shown in Figure 2a, these two potentials are symmetric with two asymmetric energy wells each. Correctly obtaining free energy difference between these two butane molecules, theoretical value of which is zero, is very challenging, because existing energy barriers are not trivial to be crossed by thermal activations. By regular canonical ensemble treatments, computed free energy difference was yielded to be 22.47 kcal/mol, which is equal to the difference between global minimum of one molecule and local minimum of the other. As shown in Figure 2, because it can simultaneously enhance sampling, DTA-SS simulation allows rough free energy value, estimated based on Equation 5, to reach the region of zero kcal/mol in 2.5 ps and evolve to be 0.2 kcal/mol with small fluctuation in about 16 ps. Finally, after totally 1 ns simulation, based on Equation 6, exact free energy is estimated to be 0.09 kcal/mol.

4 Hybrid Potential Space Random Walk to Robustly Realize QM/MM Based Simulated Scaling Simulations

The generalized ensemble algorithms can be robustly applied to the systems treated with classical energy potential functions (or called molecular mechanical (MM) potentials), which are usually described as the sum of a series of geometry-dependent functional energy terms, because in these types of potentials, energies can be robustly evaluated regardless of structural qualities, and chemical space is restrictedly pre-determined by the definition of atom types and connectivities. However, direct application of these generalized ensemble methods to the simulations treated with quantum mechanical (QM) potentials (pure QM potential or its hybrid with molecular mechanical potential (QM/MM)) can be problematic and sometimes almost impossible in commonly applied finite-time-step molecular dynamics settings, for the fact that the convergence of the self-consistent-field (SCF) calculation for QM energy evaluation is very structure-sensitive; it can be more demanding than the requirement for the ordinary differential equation propagation stability in molecular dynamics simulations. If generalized ensemble algorithms are directly applied in the QM-based (QM or QM/MM) simulations, instantaneously twisted molecular structures, inevitably generated due to high temperature (or high effective "temperature" corresponding to a low scaling parameter λ value) activations, may make electronic structural SCF calculations difficult to converge or artificially converge to other electronic structural species (effectively like the occurrence of chemical reactions). Although facing such challenge, the increasing demands of accurate calculation of free energy values urgently require robust and efficient QM-based free energy simulation methods. In order to reconcile such confliction between a necessary activation (either by increasing temperature or lowering scaling parameter λ) and the structure-sensitivity nature of QM calculations, one of possible solutions is to avoid direct walking between the activated MM states (with low λ values or high temperatures) and the state requiring QM energy and force calculations in the simulated scaling method design.

Theoretical Design of the Method. Following the same thought in one of our recent works [12], we can design the following hybrid traveling path from QM0 to MM0, then from MMo to MM1, and the from MM1 to QM1, instead of a simple path scaled by one scaling parameter as introduced in one of PI's previous developments. In this hybrid potential space, two end points are our target QM/MM states and the center path from MM0 to MM1 is the same as our classical potential based simulated scaling method.

In this scheme, we need to realize a random walk in a hybrid path rather than a single street. To do so, all the equations in the simulated scaling method can still apply, except for the expression of $\frac{\partial U}{\partial \lambda}$ in Equations 1 & 2. The values of $\frac{\partial U}{\partial \lambda}$ depend on which portion of the path the simulation is currently located in. If the simulation is in the path starting at QM0 and ending at MM0, this value should be equal to $U_{MM0} - U_{QM0}$; if the simulation is in the path starting at MM0 and ending at MM1, this value should be equal to $U_{MM1} - U_{MM0}$; if the simulation is in the path starting at MM0 and ending at MM1, this value should be equal to $U_{QM1} - U_{MM1}$. As noted earlier, Equations 1 & 2 are based on linear scaling equation. As the matter of fact, any nonlinear scaling potentials can be applied to the simulated scaling simulations. Then, these equations can be easily modified. In the original simulated scaling method implementation, soft-core potentials have been successful included, tested, and even employed for practical biomolecular studies.



(a) (b)

Figure 3. (a) Time dependent free energy difference between methanol and methane in gas phase; (b) Time dependent free energy difference between methanol and methane in the solution.

Illustrative Example. A model system is set to compute the solvation free energy difference between methoal and ethane. Here, QM/MM potential was employed; the solutes are treated with the SCCDFTB method and the solvent molecules are treated with the TIP3P model.

As shown in Figure 3, QM/MM based solvation free energy difference value can be efficiently predicted for methanol and ethane. The predicted solvation free energy difference, -7.2 kcal/mol, reproduces experimental values nicely.

References

1. D. L. Beveridge and F. M. Dicapua, Ann. Rev. Biophys. Chem. **18**, 431, 1989.
2. P. Kollman, Chem. Rev. **93**, 2395, 1993.
3. J. M. Rickman and R. LeSar, Annu. Rev. Mat. Res. **32**, 195, 2002.
4. T. Simonson, G. Archontis, and M. Karplus, Acc. Chem. Res. **35**, 430, 2002.
5. Free Energy Calculations: Theory and Applications in Chemistry and Biology, Springer Series in Chemical Physics (Eds. A. Pohorille and C. Chipot) (2007).
6. S. A. Adcock and J. A. McCammon, Chem. Rev. **106**, 1589, 2006.
7. W. Yang, H. Nymeyer, H.-X. Zhou, B. A. Berg, and R. Brüschweiler, J. Comp. Chem. **29**, 668, 2008.
8. H. Li, M. Fajer, and W. Yang, J. Chem. Phys. **126**, 024106, 2007.
9. S. Duane, A. D. Kennedy, and B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216, 1987.
10. F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050, 2001.
11. W. Kwak and U. H. E. Hansmann, Phys. Rev. Lett. **95**, 138102, 2005.
12. H. Li and W. Yang, J. Chem. Phys. **126**, 11410, 2007.

Protein-Ligand Docking Including Protein Flexibility: A Hierarchical Approach

Iris Antes, Christoph Hartmann, and Thomas Lengauer

Max-Planck-Institut für Informatik,
Campus E1 4, 66123 Saarbrücken, Germany
E-mail: antes@mpi-inf.mpg.de

To describe protein-ligand interactions realistically, it is necessary to account for the structural changes in both the receptor and the ligand during complex formation. However, in classical docking approaches the receptor is treated either rigid of semi-rigid. We developed new program, DynaDock[6], which allows full flexibility for both, the ligand and the receptor during docking. For this purpose we combined two new methods: IRECS (Iterative REduction of Conformational Space)[4] and OPMD (Optimized Potential Molecular Dynamics), which together allow an efficient sampling of ligand and receptor conformations.

1 Introduction

The major application for protein-ligand docking is the development of new drugs for efficient and safe treatment of diseases. Molecular docking methods developed for this purpose are very efficient approaches, which are specifically designed for rapid identification and characterization of protein-ligand interactions. The main application area of these methods is virtual screening. Virtual screening places tight CPU time constraints on the applied methodology due to the enormous number of potential drug molecules to be tested (around 100.000 to 1.000.000 molecules). Due to these time constraints one very rigorous approximation is commonly used: the structural changes in the receptor upon ligand binding are neglected. However, for a growing number of target structures receptor flexibility is crucial. Thus the efficient treatment of protein flexibility during docking is one of the major challenges within the field.

2 Methods

2.1 IRECS

The most commonly used methods for an efficient treatment of side chain flexibility during docking are discrete optimization approaches based on predefined side chain rotamers, like e.g. ensemble methods (e.g. FlexE[1]). Due to the large number of possible side chain conformations at the receptors binding interface, these docking approaches often face the problem of a combinatorial explosion if all possible side chain conformations are considered. Thus it is crucial to preselect a small number of flexible side chains in the binding site for which alternative conformations are used. IRECS was developed for this purpose. It is a new tool for side chain placement, which is especially tailored to the needs of molecular docking. In contrast to other side chain placement tools, which predict the same number of conformations (mostly one) to all side chains of the protein, our tool is able to predict

an ensemble of the most probable conformations for each side chain of a protein. The relative numbers of rotamers that are assigned to each side chain correspond to the side chains flexibility. The abolute level of flexibility can be defined by the user as the final rotamer density (average number of rotamers per residue in the protein) of the output structures. Thus IRECS leads to a minimal, flexibility optimized set of binding site side chain conformers. The predicted side chain ensembles can be used for ensemble based docking (FlexE). The typical application areas of IRECS/FlexE are drug design projects for which a relatively fast algorithm is required and for which a discrete approximation of side chain flexibility is sufficient.

2.2 DynaDock

Although the number of target structures for which molecular docking can be performed increases considerably by the use of discrete, rotamer based flexible docking approaches, there are several limitations of these methods for which no satisfying solution has been yet found. The main restriction is that backbone movement can be considered only to a very limited extent. Another limitation is that, to be efficient enough, a rather coarse grained definition of the conformations must be used. Thus induced fit effects upon ligand binding, which are too small to lead to new side chain conformations, but are necessary for successful binding, are not considered. Last but not least, all docking methods still have a sampling problem if the number of the rotational degrees of freedom to be considered is above a certain limit. Thus efficiently docking a very large, flexible ligand into a flexible binding site is still a challenge for the field.
We have developed a new docking algorithm, DynaDock, which is especially suited for such cases. DynaDock combines discrete search algorithms from bio- and cheminformatics with continuous biophysical simulation methods. The underlying idea behind the algorithm is to first perform a flexible ligand-rigid protein docking on the basis of a modified interaction energy function, which allows for overlapping conformations (can be performed with IRECS). These conformations are then refined by regaining the correct physical interaction energy landscape during a fully flexible simulation of the whole system. For this purpose a new simulation method was developed, Optimized Potential Molecular Dynamics (OPMD), which allows for an efficient molecular dynamics based sampling of the conformational space of both the protein and the ligand. The combined DynaDock-OPMD algorithm leads to a mutual adaptation of the 3D shapes of the ligand and the protein's binding pocket.

3 Results

The performance of IRECS was evaluated on a set of 160 crystal structures of proteins. First, final structures were predicted with a rotamer density equal one, i.e. with a single conformation per side chain. This corresponds the output of other side chain prediction programs. The results were compared to two other side-chain prediction tools, SCWRL[2] and SCAP[3]. IRECS achieved a χ_1 accuracy of 84.7% and a χ_{1+2} accuracy of 74.3%, using a 40° cutoff. The average side chain RMSD from the crystal structures was 0.78 Å. This is comparable to the performance of SCWRL[2] (χ_1 = 82.3%, χ_{1+2} = 68.0%, av. RMSD = 0.85 Å) and SCAP[3] (χ_1 = 84.0%, χ_{1+2} = 80.6%, av. RMSD = 0.82 Å).

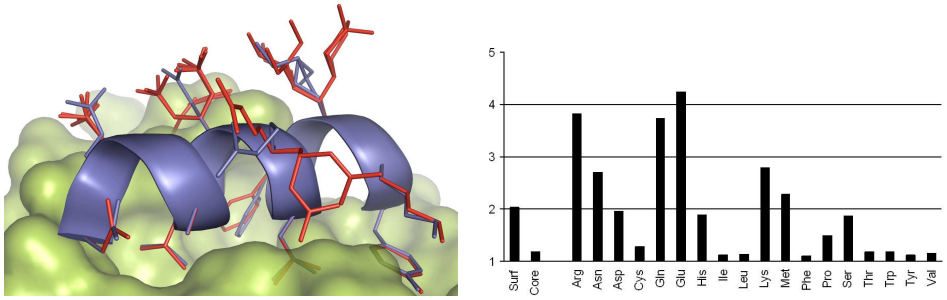


Figure 1. (a) A helix (chain B, position 69-80) of human UDP-galactose 4-epimerase (PDB: 1EK6). The side chains predicted with IRECS are red, the side chains of the crystal structure are blue. (b) Average number of rotamers assigned to the different amino acids in the final structures predicted by IRECS (rotamer density of 2).

In addition, final structures with a rotamer density of two were predicted and the average number of rotamers per amino acid type was analysed. These numbers (Figure 1b) correspond very well to the internal flexibility of the individual amino acids. Thus the ensembes of side chain rotamers assigned by IRECS are representative for the internal flexibility of individual residues.

The DynaDock method was tested on a set of 20 X-ray structures of protein-peptide complexes obtained from the PDB database. The structures were chosen to cover a wide range with respect to the peptide's length (3 to 11 residues) and its surface exposure (20 to 60 %). Starting with randomly disturbed peptide conformations around the binding site with a peptide RMSD larger than 3.5 Å, we were able to obtain refined structures with RMSD values smaller than 2.0 Å for all complexes in our test set using the OPMD approach. Analysing the RMSD of the lowest energy structures, it was observed that also these values were smaller than 2.1 Å for 11 cases and only one case was above 3.0 Å, namely 3.5 Å. These are very promising results, especially considering that also the longest peptide (11 residues) could be refined to 1.08 Å.

References

1. H. Claussen, C. Buning, M. Rarey, and T. Lengauer, *FlexE: efficient molecular docking considering protein structure variations*, J. Mol. Biol. **308(2)**, 377–395, 2001.
2. A. Adrian, A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction*, Protein Sci **12(9)**, 2001–2014, 2003.
3. Z. Xiang and B. Honig, *Extending the accuracy limits of prediction for side-chain conformations*, J. Mol. Biol. **311(2)**, 421–430, 2001.
4. C. Hartmann, I. Antes, and T. Lengauer, *IRECS: A new algorithm for the selection of most probable ensembles of side-chain conformations in protein models*, Protein Science **16(7)**, 1294–1307, 2007.
5. C. Hartmann, I. Antes, and T. Lengauer, *Virtual screening with modeled or flexible side chain conformations*, Proteins, in press.
6. I. Antes, *DynaDock: A new docking approach for protein-peptide docking including receptor flexibility*, in revision.

Electroporation Studied by Molecular Dynamics Simulations

Rainer A. Böckmann[1], Shirley W. I. Siu[1], Simon Leis[1], Bert L. de Groot[2], Sergej Kakorin[3], Eberhard Neumann[3], and Helmut Grubmüller[4]

[1] Theoretical and Computational Membrane Biology, Center for Bioinformatics Saar, 66041 Saarbrücken, Germany
E-mail: rainer@bioinformatik.uni-saarland.de

[2] Computational Biomolecular Dynamics Group, Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[3] Physical Chemistry III, Department for Chemistry, University of Bielefeld, 33501 Bielefeld, Germany

[4] Theoretical and Computational Biophysics Department, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Membrane electroporation is the method to directly transfer bioactive substances such as drugs and genes into living cells, as well as preceding electrofusion. Although much information on the microscopic mechanism has been obtained both from experiment and simulation, the existence and nature of possible intermediates is still unclear. Here, we investigated the equilibrium effect of external electric fields on the membrane structure, the effect of membrane-embedded proteins on the stability of membranes, and the kinetics of the electropore formation process, applying molecular dynamics simulations.
The results show a quadratic change of the membrane capacitance with the applied voltage, and a significant stabilization of membranes by proteins. For electropores, an average pore radius of 0.47 ± 0.15 nm was obtained, in favourable agreement with conductance measurements and electrooptical experiments of lipid vesicles. A linear dependency of the activation energy for prepore formation on the applied field is seen, with quantitative agreement between experiment and simulation. The distribution of preporation times suggests a four state pore formation model.

1 Introduction

Membrane electroporation (MEP) is nowadays an established technique to render cell membranes porous and permeable by applying electric pulses to cells in suspension, adherent cells, and tissue. Historically, the structural concept of MEP has been derived from functional changes such as cell death, the nondestructive electro-release of intracellular components from isolated organelles, and the functional direct electro-uptake of naked gene DNA into mouse lyoma cells. MEP is widely used for the efficient direct electrotransfer of all kinds of bioactive substances, in particular drugs and genes, not only in cell biology and biotechnology but also in the new medical disciplines of electrochemotherapy and electrogenetherapy used e.g. for vaccination or in RNA transfection. Other electroporative phenomena such as electrofusion of cells or electroinsertion of xenoproteins by nonpermeabilizing electroporation pulses at low voltages are intrinsically coupled to the structural changes induced by MEP.

Despite major experimental and theoretical efforts, several questions remained unanswered so far: for example, an intermediate on the way to hydrophilic pores has been proposed and tentatively been assigned a hydrophobic pore on the basis of kinetic measurements. Also, conductance measurements on planar lipid bilayers showed the existence of a nonconductive prepore state. However, in molecular dynamics (MD) simulations such a hydrophobic pore has not been observed. Accordingly, the structural features of such an intermediate are still unknown. A second issue is that no attempt has been made so far to quantitatively relate the pore formation times observed in simulations to measured pore formation kinetics, such that up to now the simulations have not been rigorously validated against experiment.

2 Methods

Bilayer patches composed of 128 to 512 POPC (DOPC) lipids surrounded by explicit water were studied. Larger systems were required for the study of equilibrium effects and of a stable electropore. Force field parameters for the lipids were taken from Berger[1] or from Siu et al.[2]. All simulations were performed at full hydration, for $5-200$ ns. Ions were considered for the study of membrane capacitance changes under external electric fields as well as for studying a stable electropore.

3 Influence of Electric Fields on Membrane Capacitance

The influence of external electric fields on the membrane capacitance was studied applying external electric fields of $0.1-0.4$ V/nm across a DOPC bilayer. In experiments, a quadratic increase of membrane capacitance with the membrane potential was described:

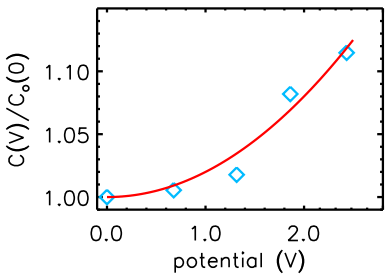


Figure 1. Membrane capacitance changes as a function of the potential drop across a lipid bilayer. The solid line displays the experimentally obtained quadratic behavior, the blue diamonds results from MD simulations applying external electric fields of varying strengths across a DOPC bilayer. The bilayer contained 256 lipids, the phospholipid force field was derived from GAFF[3].

$$\frac{C(V)}{C(0)} = 1 + \alpha V^2 \, . \tag{1}$$

Membrane capacitance changes obtained using the recently developed generalized all-atom force field based on Amber (GAFF[3]) for DOPC are in very good agreement with the experiments. We suggest this modified membrane state with an increased area per lipid and differentially tilted lipid head groups for both lipid leaflets as an intermediate state to prepore formation.

4 Influence of Gramicidin on Membrane Stability

Large potential drops across biological membranes occur already by small charge imbalances. Our results obtained in simulation studies of a DMPC/gramicidin system suggest that gramicidin considerably suppresses membrane electropore formation[4]. Pore formation times were increased at least by a factor of three.

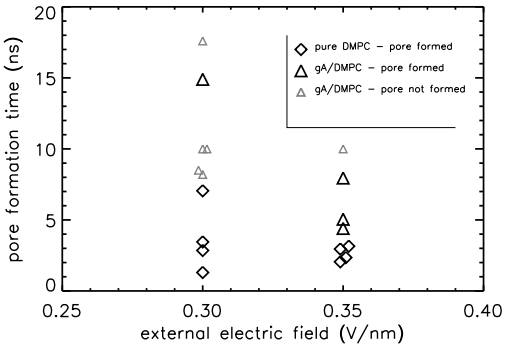


Figure 2. Pore formation times observed in MD simulations for a pure DMPC lipid bilayer (black diamonds) and for a mixed gramicidin/DMPC system (black triangles, gramicidin in DH conformation) at two different field strengths (0.3 and 0.35 V/nm). Small gray symbols denote the simulation lengths of further simulations in which pore formation could not be observed. The total simulation time exceeds 120 ns.

5 Kinetics of Electropore Formation

Comparison of prepore formation rates obtained from 50 simulations at varying field strengths with experiment allowed to estimate the average number of lipids involved in electropore formation to 140. From this result one would predict that pores are separated typically by approximately 7 nm. The radius of stable electropores was estimated to 0.47 ± 0.15 nm from a simulation with properly adjusted external electric field[5]. Pore formation is preceded by two intermediate steps: (1) tilting of lipid headgroups, coupled to an increase in the lipid area; (2) prepore intermediate involving $2-5$ lipids[5].

Acknowledgments

Financial support by the DFG (Graduate School *Structure Formation and Transport in Complex Systems*, No. 1276/1) is acknowledged (SWIS and RAB). As member of the Center for Bioinformatics, RAB is supported by the DFG Grant BIZ 4/1.

References

1. O. Berger, O. Edholm, and F. Jähnig, *Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.*, Biophys. J. **72**, 2002–2-13, 1997.
2. S. W. I. Siu, R. Vacha, P. Jungwirth, and R. A. Böckmann, *Biomolecular simulations of membranes: Physical properties from different force fields.*, J. Chem. Phys. **128**, 125103, 2008.

3. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *Development and testing of a general AMBER force field*, J. Comp. Chem. **25**, 1157–1174, 2004.
4. S. W. I. Siu, and R. A. Böckmann, *Electric Field Effects on Membranes: Gramicidin A as a Test Ground*, J. Struct. Biol. **157**, 545–556, 2007.
5. R. A. Böckmann, B. L. de Groot, S. Kakorin, E. Neumann, and H. Grubmüller, *Kinetics, Statistics, and Energetics of Lipid Membrane Electroporation Studied by Molecular Dynamics Simulations.*, Biophys. J., *in press*, 2008.

Modeling Conformational and Molecular Weight Heterogeneity with Analytical Ultracentrifugation Experiments

Borries Demeler[1] **and Emre Brookes**[2]

[1] The University of Texas Health Science Center at San Antonio,
Department of Biochemistry, San Antonio, Texas, USA
E-mail: demeler@biochem.uthscsa.edu

[2] The University of Texas at San Antonio,
Department of Computer Science, San Antonio, Texas, USA
E-mail: ebrookes@cs.utsa.edu

Sedimentation velocity experiments reveal information about molecular weight and shape of sedimenting macromolecules. The observables in such experiments are the sedimentation and diffusion coefficients and the concentration of individual solutes. We have developed parallel optimization algorithms that allow us to extract molecular parameters from mixtures of macromolecules using a nearly model-independent approach. Using a combination of deterministic and stochastic optimization, we are able to fit complex analytical ultracentrifugation experiments globally with excellent convergence properties. Our software uses the TIGRE grid middleware to distribute the computing effort to Teragrid and other computing resources, and offers a public web portal for the hydrodynamic analysis of AUC experiments[1]. Our solutions provide unparalleled resolution, and allow us to characterize polymerization events, aggregation and provide high resolution information in structure and function studies in the solution state.

1 Introduction

The sedimentation and diffusion transport of a solute observed in an analytical ultracentrifugation (AUC) sedimentation velocity experiment is described by the Lamm equation[2]. Mixtures of solutes can be modeled well by a linear combination of finite element solutions of the Lamm equation[3,4] where each term represents a different solute in the mixture. The sedimentation (s_k) and diffusion coefficients (D_k) are parameters of the Lamm equation, and define uniquely the molecular weight and shape of each solute k in the mixture, while the amplitude of each term determines the partial concentration (c_k). In an AUC experiment the goal is to correctly determine s, D, c as well as n, the number of solutes present in the mixture. The inverse problem of fitting experimental data to simulations of Lamm equation systems represents a difficult optimization problem which is nonlinear with respect to the fitting parameters. We present here a method for evaluating experimental data by applying multiple optimization algorithms in series for obtaining the most likely parsimonious parameter distribution that satisfies Occam's razor. Our approach is implemented on a parallel computing platform utilizing the globus-based TIGRE grid middleware[5] which can be conveniently accessed through a web portal. Results can be further analyzed with the UltraScan software[6,7]. Our approach includes algorithms for initialization, systematic noise deconvolution, parameter search and parsimonious regularization.

2 Initialization

The parameter search requires an initialization step which identifies the limits of the domains of two of the fitting parameters, s, and D. If the experimental data contain significant amount of time invariant systematic noise, the s limits are conveniently identified with the time-derivative method by Stafford, which yields a model-free transformation of the primary data that eliminates any time invariant noise contributions[8]. A more accurate initialization can be obtained from the experimental data by the enhanced van Holde - Weischet method[9], which yields a model-independent, diffusion-corrected sedimentation distribution for cases where time invariant noise is not significant. Accurate limits for D are difficult to obtain reliably by model-independent means, and require prior knowledge and parameterization by the frictional ratio, f_r:

$$D_k = \frac{RT}{18\pi N}\left[\frac{2(1-\bar{\nu}_k\rho)}{s\,\bar{\nu}}\right]^{\frac{1}{2}}(\eta f_{r,k})^{-\frac{2}{3}} \tag{1}$$

where R is the gas constant, T the temperature in Kelvin, N is Avogadro's number, η and ρ are the viscosity and density of the solvent, and $\bar{\nu}_k$ is the partial specific volume of solute k. Values for f_r are chosen based on the analyte under investigation, for example 1-2 for globular macromolecules, 1.5-3 for disordered or denatured proteins, or values up to 10 for elongated molecules such as long nucleic acids, fibrils or amyloid aggregates. For unknown systems a sufficiently large range can also be chosen, but in those cases additional refinement steps may be required.

3 Time-Invariant Noise Reduction and 2-Dimensional Spectrum Analysis Parameter Search

The precision of parameter estimation is inversely correlated with the experimental noise present in the primary data. It is therefore important that systematic noise contributions resulting from instrument flaws are accounted for and that stochastic noise contributions are attenuated using Monte Carlo (MC) methods[10]. We have shown that systematic noise contributions can be effectively eliminated using algebraic means[11]. Experimental design considerations can further improve noise characteristics, for example, by using intensity measurements instead of absorbance measurements, stochastic noise is reduced by a factor of $\approx \sqrt{2}$ by not subtracting the reference signal. This subtraction leads to the convolution of two stochastic noise vectors and an increase in the stochastic noise. In the first optimization step we perform a 2-dimensional spectrum analysis between the limits determined above in section 2 as described by Brookes et al.[12]. Briefly, a divide and conquer algorithm is employed to search multiple coarse-grained subgrids spanning the entire 2-dimensional parameter range in s and f_r. Each grid point is an element in a linear combination of finite element solutions of the Lamm equation, whose amplitudes are determined in a least squares fit using the non-negatively constrained linear least squares (NNLS) fitting algorithm[13]. By combining the results from multiple relatively coarse grids that are slightly offset against each other, a high-resolution, 2-dimensional spectrum analysis (2DSA) is obtained. The result is a sparse matrix identifying potential signals in the sample.

4 Parsimonious Regularization of the 2DSA Grid Using Genetic Algorithms

After performing the 2DSA analysis, a sparse grid identifying potential solutes is obtained. However, due to the presence of experimental noise and due to the degeneracy resulting from fitting with an overdetermined system the result is subject to the presence of false positives. While such effects cannot be entirely eliminated, a parsimonious regularization can improve the solution significantly by providing a solution that satisfies Occam's razor. Occam's razor states that the solution with the greatest parsimony of parameters resulting in nearly the same residual mean square deviation (RMSD) as another less parsimonious solution is to be preferred. We have implemented a second step in the optimization process which takes advantage of a genetic algorithm (GA) approach. In this approach, we initialize the GA analysis with parameter constraints obtained by drawing 2-dimensional boundaries with user-defined width around each solute. Overlaps between adjacent boxes are eliminated by further subdividing existing boxes to create new, non-overlapping boxes. During fitting, parameters are adjusted in an evolutionary approach based on fitness:

$$fitness = RMSD * \left(1 + (rf * nz(x))^2\right) \quad (2)$$

where nz is the cardinality of solution x and rf is a regularization factor applying RMSD penalties to increase parsimony[14].

5 Global Multi-Speed Genetic Algorithm Monte Carlo Refinement

In order to enhance the information content of AUC experiments, data from multiple experiments performed at different speeds can be combined in a global fit. In high speed experiments, sedimentation signals are enhanced, in low speed experiments diffusion signals are improved. GA-MC analysis can be performed globally by constraining the solute model to all datasets. Table 1 lists results from a simulated 5-component system with heterogeneity in both shape and molecular weight (realistic noise added), representing a linear elongation event, performed at both 20 krpm and 60 krpm.

Solute	Molecular Weight (kD)	Partial Concentration	Frictional Ratio, f/f0
1	24.26 (24.20, 24.33) [25]	0.0972 (0.0966, 0.0982) [0.1]	1.21 (1.21, 1.21) [1.2]
2	48.04 (47.74, 48.46) [50]	0.102 (0.101, 0.104) [0.1]	1.41 (1.40, 1.42) [1.4]
3	100.2 (97.96, 101.8) [100]	0.0995 (0.0982, 0.101) [0.1]	1.65 (1.63, 1.67) [1.6]
4	198.0 (194.2, 200.8) [200]	0.0996 (0.0989, 0.101) [0.1]	1.84 (1.82, 1.86) [1.8]
5	385.3 (380.4, 394.0) [400]	0.100 (0.100, 0.101) [0.1]	2.01 (1.99, 2.04) [2.0]

Table 1. GA-MC results for a global fit of a multispeed 20/60 krpm experiment (described in text). The results demonstrate remarkable agreement with the target. Parentheses: 95% confidence intervals; square brackets: target value. All values rounded off to 3 or 4 significant digits.

Acknowledgments

We would like to thank the NIH-NCRR for funding this research through Grant RR022200, and the NSF for Teragrid allocation TGMCB070038 (both to B.D.).

References

1. Demeler, B. 2006. *A web-based Laboratory Information Management System for UltraScan*. `http://www.uslims.uthscsa.edu`.
2. Lamm, O., 1929. *Die Differentialgleichung der Ultrazentrifugierung*. Ark. Mat. Astron. Fys. **21B**, 1–4, 1928.
3. Cao, W. and B. Demeler. *Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation*. Biophys. J. **89**, 1589–1602, 2005.
4. Cao, W. and B. Demeler. *Modeling Analytical Ultracentrifugation Experiments with an Adaptive Space-Time Finite Element Solution for Multi-Component Reacting Systems*. Biophys. J. **95(1)**, 54–65, 2008.
5. The Texas Internet Grid for Research and Education. A Globus-based grid middleware for sharing computational resources across a wide area network, `http://www.tigre.net`.
6. Demeler, B. (2008) UltraScan: A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. Version 9.7, `http://www.ultrascan.uthscsa.edu`.
7. Demeler, B. (2005) *UltraScan A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments*, in Modern Analytical Ultracentrifugation: Techniques and Methods (D. J. Scott, S. E. Harding, and A. J. Rowe, Eds.) 210-229, Royal Society of Chemistry (UK).
8. Stafford, W. F. *Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile*. Anal. Biochem. **203**, 1–7, 1992.
9. Demeler, B. and K. E. van Holde *Sedimentation velocity analysis of highly heterogeneous systems*. Anal Biochem **335(2)**, 279–88, 2004.
10. Demeler, B. and E. Brookes. *Monte Carlo analysis of sedimentation experiments*. Colloid Polym Sci **286(2)**, 129-137, 2008.
11. Schuck P and Demeler B. *Direct sedimentation analysis of interference optical data in analytical ultracentrifugation*. Biophys J. **76(4)**, 2288-96, 1999.
12. Brookes, E. H., R. V. Boppana and B. Demeler. *Computing Large Sparse Multivariate Optimization Problems with an Application in Biophysics.* SuperComputing 2006 Conference Proceedings 0-7695-2700-0/06 2006 IEEE (`http://sc06.supercomputing.org/schedule/pdf/pap320.pdf`).
13. Lawson, C. L. and Hanson, R. J. 1974. *Solving Least Squares Problems.* Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
14. Brookes, E and B. Demeler. *Parsimonious Regularization using Genetic Algorithms Applied to the Analysis of Analytical Ultracentrifugation Experiments*. GECCO Proceedings ACM 978-1-59593-697-4/07/0007 (2007).

Exploring the Free Energy Surface of Short Peptides by Using Metadynamics

Carlo Camilloni[1] **and Alfonso De Simone**[2]

[1] Department of Physics and INFN, University of Milano,
via Celoria 16, Milan 20133, Italy
E-mail: carlo.camilloni@mi.infn.it

[2] Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
E-mail: ad491@cam.ac.uk

The free energy surface (FES) of three Poly-Alanine peptides is exhaustively reconstructed by using metadynamics. A specific collective coordinate has been introduced to account for the hydration of the N- terminal region of alpha-helical conformations. Calculations suggest that Poly-Alanine peptides mainly populate unfolded states. Interestingly FESs exhibit different properties among peptides thereby providing some hints on factors determining the conformational preferences of short Poly-Alanine sequences. Overall the calculations evidence the efficiency of Metadynamics in exploring free energy surfaces of biological relevant molecules.

1 Introduction

Biological macromolecules have typically access to a wide spectrum of conformations. Such a dynamical behavior, essential for solving the biological functions, is connected to a complex free energy surface (FES). With the growing power of calculators and the continuous refinement of force fields, molecular simulations have been increasingly employed to address for relevant biological topics. In this scenario, a demand for innovative tools allowing for an efficient exploration of the FES has pushed for the development of a variety of novel methods. Herein, to reconstruct the FES of polypeptides, we employ a method - Metadynamics - recently introduced by Parrinello and coworkers[1] and successfully applied to a large variety of scientific problems from physics to chemistry and biology. As test case, the FESs of small Poly-Alanine peptides are sampled.

2 Methods

Metadynamics performs an efficient exploration of multidimensional FES by means of collective coordinates (CVs) and a history dependent potential (eq.1):

$$F_G(s,t) = \int_0^t dt' W \exp\left(-\frac{s - s(x(t'))^2}{2\delta^2}\right) \quad (1)$$

The dynamics in the space of the CVs is guided by the free energy of the system plus the history-dependent potential which sums Gaussians of width d and weight W centered along the CVs trajectory up to time t (for further details see[1]). Herein, we employed three different CVs. The first is the gyration radius calculated on the C-alpha atoms of the peptide. The second is the root mean square deviations of the dihedral angles compared to an

ideal alpha-helix. Finally, to account for the conformational stability of helical conformations, we biased the hydrogen bonds that the 4 N-terminal residues make with water (see discussion). The latter CV is accounted with the following formula:

$$\sum_{i=1}^{4} \sum_{j=1}^{nr.Waters} \frac{1-(\frac{r_i-r_j}{d_0})^6}{1-(\frac{r_i-r_j}{d_0})^{12}} \tag{2}$$

with $d_0 = 0.22$ nm and the *i* running on the amide hydrogen atoms of the first four residues of the peptide (representing the N-terminal loop of an ideal alpha-helix) and the *j* running on all the acceptor oxygen atoms of the solvent. We simulated three poly-Alanine peptides of length 9, 14 and 19 (PolyA$_9$, PolyA$_{14}$ and PolyA$_{19}$, respectively). The simulations have been performed by using GROMETA[2,3], which is a metadynamics module developed to run in association with GROMACS, and GROMOS 53A6 forcefield. The system has been accommodated in a dodecahedron box with periodic boundary conditions. The box has been filled with SPCE water molecules. Peptide N- and C- terminal moieties have been modeled as uncharged. Non-bounded interactions have been treated by using a cutoff for Van der Waals (1.4nm) and PME for long-range electrostatic interactions (mesh spacing 0.125 nm). The dynamics have been simulated in the canonical ensemble, coupled with a Nose-Hoover thermal bath. The equilibration procedure has been performed as in Ref. 2. For each peptide, simulations have been carried out for 60ns starting by an ideal alpha-helical conformation.

3 Results and Discussion

We performed exhaustive metadynamics runs of PolyA$_9$, PolyA$_{14}$ and PolyA$_{19}$ with particular focus on the N-terminal hydration, for which a specific CV has been written. Indeed, the latter is known to be one of the possible strategies that nature adopts to stabilize alpha-helices[4]. To improve the sampling convergence in meaningful regions of the FES, the weight W has been rescaled according to the Well-Tempered dynamics[1]. This allowed us to monitor convergence of the samplings. The 3D FES of PolyA$_9$ (Fig. 1) evidence that this peptide is unable to sample alpha-helical conformations. Indeed the peptide essentially populates unfolded states featured by a large distribution of gyration radiuses. The sampling evidences a minimally populated beta-hairpin state and an ensemble of kinked states, mainly featured by a central Hbond promoting a beta-turn. This result is in a good agreement with a recent NMR study[5]. On the other hand, the larger PolyA$_{14}$ (Fig. 2) and PolyA$_{19}$ (not shown) are able to marginally explore alpha helical regions. Notably, by passing from 9 to 14 and 19 Alanine residues, the distribution of gyration radiuses is dramatically affected with longer chains showing preference for collapsed structures (see 1D projections on CV2 in Fig. 1 and 2). It is very likely that the tendency to adopt more collapsed structures is a first step toward the formation of regular secondary elements as shown by the mild exploration of alpha-helix and a larger population of beta-hairpins. It is worth nothing that it was not possible to fully test the newly introduced N-terminal hydration CV since the simulated peptides are not properly able to populate the alpha helical structures. However, it is likely that for more helical peptides this CV would efficiently help the convergence of the Metadynamics sampling. Future development will focus on alpha-helical peptides for which a detailed thermodynamics and kinetics description is available in literature.

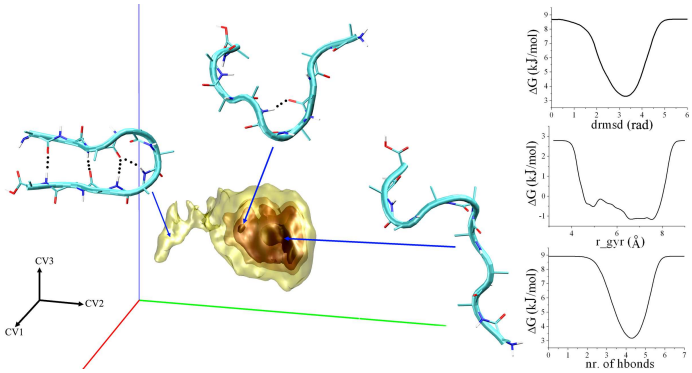


Figure 1. **FES of PolyA$_9$.** The left plot reports 3D FES (kJ/mol) with curves contoured ad (-10, -11, -11.8 and -12.4 kJ/mol). Blue arrows connect representative structures with respective regions of FES. Right panel reports FES projected on CV1 (top), CV2 (middle) and CV3 (bottom).

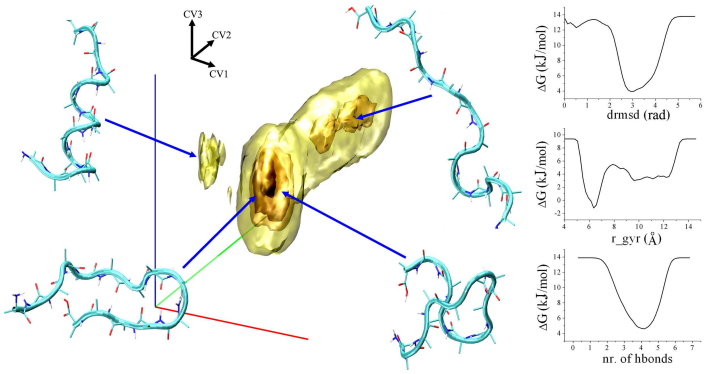


Figure 2. **FES of PolyA$_{14}$.** The plot is structured as in figure 1. Contour curves of the 3D FES are plotted at -7, -14, -16 and -18 kJ/mol.

Acknowledgments

A.D. is supported by European Molecular Biology Organization.

References

1. (a) A. Laio, M. Parrinello PNAS **99**, 12562-12566, 2002.
   (b) A. Barducci, G. Bussi, M. Parrinello Phys. Rev. Lett. **100**, 020603, 2008.
2. C. Camilloni, D. Provasi, G. Tiana, R. A. Broglia, Proteins **71**, 1647-1654, 2008.
3. C. Camilloni, D. Provasi, G. Tiana, R. A. Broglia, *Gromacs Metadynamics Package* `http://lxmi.mi.infn.it/~provasi/grometa/Site/Welcome.html`.
4. Y. Harpaz, N. Elmasry, A. R. Fersht, K. Henrick, PNAS **91**, 311-315, 1994.
5. J. Graf, P. H. Nguyen, G. Stock, H. Schwalbe JACS **129**, 1179-1189, 2007.

Association of Neamine and its Derivative with the Ribosomal A-Site RNA

Maciej Długosz and Joanna Trylska

Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw,
Żwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: {M.Dlugosz, J.Trylska}@icm.edu.pl

With Brownian dynamics (BD) we study the mechanism and kinetics of association of neamine, an aminoglycosidic antibiotic, and its proposed derivative with the ribosomal A-site RNA. We compare the performance of different antitibiotic models used during BD simulations.

1 Introduction

The aminoacyl-tRNA site (A-site) in the 30S bacterial ribosomal subunit is the binding site for most aminoglycosidic antibiotics[1]. Binding of aminoglycosides interferes with translation and causes the decrease of its fidelity[2]. Although used in medical therapy, antibiotics from this family suffer from moderate affinity, inadequate specificity and may cause damage to mammalian and kidney cells. Also, bacterial resistance limits their effectiveness in medical therapy. Therefore, different theoretical[3–6] and experimental[3,7,8] approaches are applied in order to understand their binding mechanism and efforts to improve their selectivity and efficiency are being made. We believe that for aminoglycosides' inhibitory role, it is not only important how strong are the bound complexes (which is usually one of the criteria applied during computer-aided drug design process) but also how fast they can be formed. Therefore, we applied the BD technique to examine first stages of binding of the neamine and designed by us model compound (Figure 1). Neamine derivative was proposed during a two stage screening procedure which utilized a pharmacophoric search for possible binding modes, followed by a binding affinity estimation (Piotr Setny, unpublished results). The antibiotic target used in BD simulations is a symmetric RNA oligonucleotide containing two ribosomal A-sites[9] (Figure 1). In our recent work[4] we studied association of various aminoglycosides with this RNA fragment. In this study improved models of antibiotics are used in an attempt to validate previous results.

2 Methods

2.1 Brownian Dynamics Methodology

BD allows one to simulate the diffusional motion between interacting solutes and compute diffusion limited rate constants of their association. The ligand is represented as a polymer composed of spherically symmetric subunits, with centrally assigned partial charges, diffusing through the electrostatic field generated by a receptor. Ligand's motion can be derived using Ermak-McCammon[10] propagation scheme:

$$r_i^{n+1} = r_i^n + \sum_j \frac{\Delta t}{k_B T} D_{ij}^n F_j^n + R_i(\Delta t) \qquad (1)$$

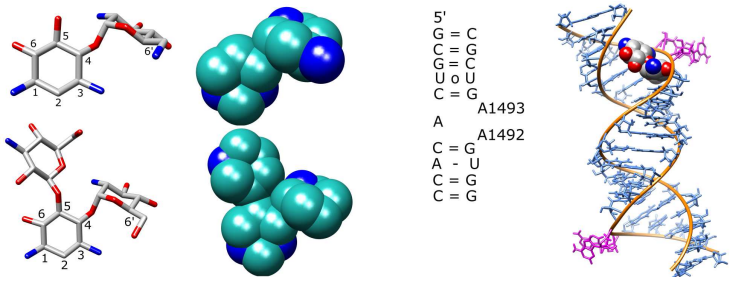


Figure 1. Left: Atomic structures and bead models used in BD of neamine (top) and its derivative (bottom). Blue beads correspond to amine groups and bear charges $+1$e. Green beads are neutral. Right: A-site RNA duplex. Secondary structure of the asymmetric unit (containing one A-site). Adenines 1492 and 1493 (*E. Coli* numbering) are displaced upon binding of an antibiotic[2]. Base pairs are represented with $=$ and $-$ (corresponding to three or two hydrogen bonds, respectively) for Watson-Crick pairs and o for non-Watson-Crick pairs. Three dimensional structure of the RNA duplex (blue and orange) with the A-site occupied by neamine (shown as van der Waals spheres); A1492 and A1493 are denoted in magenta.

where indices i and j run over coordinates of N subunits $(1 \leq i, j \leq 3N)$, r_i is the position vector component, F_i is the sum of intersubunit and external forces acting in direction i, integer n represents discrete times $t = n\Delta t$ at intervals Δt, D_{ij} is the diffusion tensor, and $R_i(\Delta t)$ is a random vector whose average is zero and $< R_i(\Delta t)R_j(\Delta t) >= 2D_{ij}^n\Delta t$. In BD, a ligand moves in the potential generated by fixed, rigid acceptor, and obtained from the solution of the Poisson-Boltzmann equation[11] on the 3D grid. External forces acting on the ligand are computed as a sum of exclusion and electrostatic terms. To estimate association rates, a large number of BD trajectories is generated. Each trajectory begins with the ligand placed randomly on the surface of the sphere with radius b. In a simulation, the ligand either diffuses outside the sphere with radius q (where $q >> b$) and the trajectory is truncated or forms an encounter complex with the receptor, satisfying some predefined reaction criteria. The ratio of the number of reactive trajectories to their total number allows one to estimate the association rate constant[12].

2.2 Brownian Dynamics Simulations Setup

The coordinates of the symmetric RNA fragment (Figure 1) with two paromomycins bound at A-sites were taken from the Protein Data Bank (code 1J7T[9]). Paromomycin coordinates were used as a template for positioning the neamine and model compound within the A-site and to establish reaction criteria defining the encounter complex. We represented neamine with 13 and its derivative with 18 spherical subunits (beads). Beads were centered on atoms of antibiotic's rings (Figure 1). Aminoglycosides were shown to bind to RNA in their fully protonated state[13] thus we assigned a total net charge of $+4$e to both molecules (distributions of partial charges are shown in Figure 1). Ligand and RNA hydrodynamic parameters were derived from their all-atom structures using the procedure described previously[4]. Resulting translational diffusion coefficients are $4.1 \cdot 10^{-6} \frac{cm^2}{s}$ for neamine and $3.5 \cdot 10^{-6} \frac{cm^2}{s}$ for its derivative. Electrostatic calculations and BD simulations utilized the University of Houston Brownian Dynamics (UHBD)[14] and are described in Ref. 4.

3 Results

	rate constant $\pm$ error $[10^{10}\frac{1}{Ms}]$	
I[mM]	neamine	model compound
150	2.42 ± 0.03	1.87 ± 0.03
200	2.22 ± 0.04	1.67 ± 0.03
250	2.10 ± 0.04	1.48 ± 0.03
300	1.86 ± 0.04	1.25 ± 0.03

Table 1. Association rate constants and their dependence on ionic strength derived from BD simulations. Error values are estimated at the 90% confidence level.

Rates of association and their dependence on ionic strength are shown in Table 1. Neamine, which is smaller than its derivative diffuses faster. For both antibiotics a rather weak dependence of computed association rates on ionic strength is observed. The decrease in association rates upon change of ionic strength is about 23% for neamine and about 33% for its proposed derivative. Both molecules bear the same net total charge and the observed difference arises from different distributions of partial charges (as the total charge of neamine is distributed in smaller volume).

In our previous study[4], we represented neamine with a much simpler, two-bead model. The observed decrease in association rate upon changing ionic strength from 150mM to 300mM was about 5% and computed association rates had slightly higher values. A more detailed 13-bead model used in the present study probably better reproduces the effects arising from the distribution of partial charges, thus larger changes in neamine's association rates with ionic strength. Nevertheless, the influence of ionic strength on the binding kinetics is rather insignificant in both cases.

Previously, we observed that aminoglycosidic antibiotics slide along the RNA groove prior to binding. Simulations with detailed antibiotic models for neamine and its derivative reveal similar behavior.

4 Conclusions

We applied BD to determine the association rate constants of neamine and its proposed derivative. The compounds were modeled with an increased number of beads in comparison to our previous study[4]. Results obtained by describing aminoglycosides with just a few and over 10 beads showed comparable mechanisms of the encounter complex formation, similar magnitude of association rate constants but stronger dependence on ionic strength with a more detailed description of antibiotic.

Acknowledgments

We would like to thank Piotr Setny for providing us the atomic model and parameters of the neamine derivative. The authors acknowledge support from University of Warsaw

(115/30/E-343/S/2007ICM BST1255, 115/30/E-343/BST1345/ICM2008, G31-4), Polish Ministry of Science and Higher Education (3 T11F 005 30), Fogarty International Center (R03 TW07318) and Foundation for Polish Science.

References

1. F. Walter, Q. Vicens, E. Westhof, *Aminoglycoside-RNA interactions*, Curr. Opin. Chem. Biol. **3**, 694–704, 1999.
2. J. M. Ogle, V. Ramakrishnan, *Structural insights into translational fidelity*, Annu. Rev. Biochem. **74**, 129–177, 2005.
3. G. Yang, J. Trylska, Y. Tor, J. A. McCammon, *Binding of aminoglycosidic antibiotics to the oligonucleotide A-site model and 30S ribosomal subunit; Poisson-Boltzmann model, thermal denaturation, and fluorescence studies*, J. Med. Chem **49**, 5478–5490, 2006.
4. M. Dlugosz, J. Antosiewicz, J. Trylska, *Association of aminoglycosidic antibiotics with the ribosomal A-site studied with Brownian dynamics*, J. Chem. Theory Comput. **4**, 549–559, 2008.
5. T. Hermann, E. Westhof, *Docking of cationic antibiotics to negatively charged pockets in RNA folds*, J. Med. Chem. **42**, 1250–1261, 1999.
6. A. C. Vaiana, E. Westhof, P. Auffinger, *A molecular dynamics study of an aminoglycoside/A-site RNA complex: conformational and hydratation patterns*, Biochimie **88**, 1061–1073, 2006.
7. D. -H. Ryu, R. R. Rando, *Aminoglycoside binding to human and bacterial A-site rRNA decoding region constructs*, Bioorg. Med. Chem. **9**, 2601–2608, 2001.
8. P. Pfister, S. Hobbie, C. Brüll, N. Corti, A. Vassela, E. Westhof, E. C. Böttger, *Mutagenesis of 16S rRNA C1409-G1491 base-pair differentiates between 6'OH and 6'NH_3^+ aminoglycosides*, J. Mol. Biol. **346**, 467–475, 2005.
9. Q. Vicens, E, Westhof, *Crystal structure of paromomycin docked into eubacterial ribosomal decoding A-site*, Structure **9**, 647–658, 2001.
10. D. L. Ermak, J. A. McCammon, *Brownian dynamics with hydrodynamic interactions*, J. Chem. Phys **69**, 1352–1360, 1978.
11. M. K. Gilson, B. Honig, *Calculating electrostatic interactions in biomolecules: method and error assessment*, J. Comput. Chem. **9**, 327–335, 1987.
12. S. H. Northrup, S. A. Allison, J. A. McCammon, *Brownian dynamics simulations of diffusion influenced bimolecular reactions*, J. Chem. Phys. **80**, 1517–1524, 1984.
13. C. M. Barbieri, A. R. Srinivasan, D. S. Pilch, *Deciphering the origins of observed heat capacity changes for aminoglycoside binding to procaryotic and eucaryotic ribosomal RNA A-sites: a calorimetric, computational, and osmotic stress study*, J. Am. Chem. Soc. **126**, 14380–14388, 2004.
14. J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, J. A. McCammon, *Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program*, Comput. Phys. Commun. **91**, 57–95, 1995.

Simplified Approaches to Complex Biological Systems

Nikolay V. Dokholyan

Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC 27599, USA
E-mail: dokh@med.unc.edu

Some of the emerging goals in modern medicine are to uncover the molecular origins of human diseases, and ultimately contribute to the development of new therapeutic strategies to rationally abate disease. Of immediate interests are the roles of molecular structure and dynamics in certain cellular processes leading to human diseases and the ability to rationally manipulate these processes. Despite recent revolutionary advances in experimental methodologies, we are still limited in our ability to sample and decipher the structural and dynamic aspects of single molecules that are critical for their biological function. Thus, there is a crucial need for new and unorthodox techniques to uncover the fundamentals of molecular structure and interactions. We follow a hypothesis-driven approach, which is based on tailoring simplified protein models to the systems of interest. Such an approach allows significantly extending the length and time scales for studies of complex biological systems. Here we describe several recent studies that signify the predictive power of simplified protein models within the hypothesis-driven modeling approach utilizing rapid Discrete Molecular Dynamics (DMD) simulations.

1 Introduction

Despite the multiple innovations in the field of molecular simulations, the size of biological molecules and complexes and the time scales at which they function remain unreachable to traditional computational approaches[1]. This roadblock hallmarks the principal challenge in computational structural biology and is the subject of our current research.

To circumvent the problem with reaching biologically-relevant time and length scales, one must simplify a biological system to elements, essential to a regime of interest. For example, if we are interested in large-scale motions of proteins occurring at the milliseconds to seconds time scale (e.g. protein folding), it is often safe to eliminate atomic vibrations occurring at time scales of picoseconds. Such time scale decoupling is an important maneuver that has been utilized in molecular dynamics simulations. However, despite such innovations as world-wide distributed computing and hardware-customized molecular dynamics, the time scales reachable by these sophisticated techniques are still limited by microseconds time for relatively small systems.

Developed in 1959 approach Discrete Molecular Dynamics (DMD)[2] has a philosophically different event-handling scheme, which makes it extremely efficient. The algorithm is based on satisfying the same basic physical principles as traditional molecular dynamics. However, instead of integrating equations of motions (integrating time $\sim N^3$, where N is the number of particles), DMD engine searches for collision events (search time $\sim N \log N$). This difference makes DMD speed by far more superior to traditional algorithms. With developed all atom force field Medusa[3], DMD simulations allow to extend visible simulation range to microseconds simulations. Next, we will describe applications of DMD engine and Medusa force field to studies of protein folding and modeling, and demonstrate various biotechnological applications.

2 Protein Design and Evolution

Studies of known proteins have revealed the intriguing co-organization of their sequences and structures. Proteins with sequence similarity higher than 25-30% usually adopt a similar structure and are called homologs, while those with low sequence similarity ($<$20%) can share the same structure and are referred as analogs3. The origin of such co-organization has been the topic of extensive discussions in the protein folding, design and evolution communities, since an understanding of the emergence of homologs and analogs in the protein universe has broad implications on our ability to rationally manipulate proteins. We developed a molecular modeling and design method, Medusa, to computationally design diversified protein sequences that correspond to similar backbone structures. It is these backbone structures that determine protein fold family. Using protein design, we directly demonstrated the formation of distinct protein homologs within a specific fold family when the structure deviates only 1-2 Å away from the original structure[3]. The study suggests that subtle structural changes, which appear due to accumulating mutations in the families of homologs, lead to a distinct packing of the protein core and, thus, novel compositions of core residues. The latter process leads to the formation of distinct families of homologs. This work was important in two ways: Firstly, we demonstrated that using protein design we could mimic the formation of the extremely complex protein universe. Secondly, we developed a new force field that enables us to design proteins and predict protein structure when coupled with the rapid DMD simulations technique.

3 Protein Design and Biotechnology

Mutagenesis is a central tool of molecular biology, genetics, and biotechnology. Knowing to what extent mutations affect the thermodynamic stability and structures of proteins is often vital for designing experiments. Estimation of protein stabilities remains a paramount challenge in computational molecular biology. We extended Medusa to a novel methodology, Eris (`http://eris.dokhlab.org`), for accurately predicting the mutation-induced protein stability changes[4,5]. Due to the complex nature of the interactions involved in protein folding, existing stability prediction methods often use empirical parameters trained on experimental protein stability data, which makes these methods highly dependent upon the training databases. Limited by their capability to model the structural changes induced by mutations, the applications of the developed methods are often restricted to mutations from large residues to small ones. We addressed these deficiencies with a unique approach that combines a physical force field and a fast conformation-sampling algorithm in an atomic framework of proteins. We showed that Eris could effectively detect and resolve the atomic clashes and structure strains introduced by the mutations and yield reliable predictions of the stability change for these mutants. We expect Eris, which is freely accessible through our server, to have applications on a broad range of mutations in the course of protein engineering. In fact, we have already used Eris to design a peptide to a cysteine-rich intestinal protein one (CRIP1), which has been identified as a novel marker for early detection of cancers. The designed peptide binds to CRIP1 at $K_d \sim 3\mu M$ *in vivo*[6]. Our all atom DMD models now feature Medusa force field, which offers a new automated technology for rational protein engineering and conformational exploration.

4 Protein Modeling

The deletion of the single residue Phe-508 in CFTR present in ∼90% of cystic fibrosis patients prevents maturation of CFTR in endoplasmic reticulum. While this mutation does not significantly affects the structure and thermodynamics of the nucleotide binding domain's (NBD1) where it resides, clearly Phe-508 deletion disrupts tertiary interactions with other domains. The lack of CFTR's 3D structure hampers our understanding how mutation affects channel's both tertiary organization and function. Using homology modeling and DMD/Medusa simulations, we constructed a three-dimensional model of CFTR structure[7]. We have further validated the structure *in vitro* by testing the residues' proximities in our model via engineered disulfide bonds between tested residues. We found that Phe-508 mediates a tertiary interaction between the surface of NBD1 and a cytoplasmic loop (CL4) in the C-terminal membrane-spanning domain. This crucial cytoplasmic membrane interface, which is dynamically involved in regulation of channel gating, explains the known sensitivity of CFTR assembly to many disease-associated mutations in CL4 as well as NBD1 and provides a sharply focused target for small molecules to treat cystic fibrosis. In addition to identifying a key intramolecular site to be repaired therapeutically, our findings advance understanding of CFTR structure and function and provide a platform for focused biochemical studies of other features of this unique ATP-binding cassette (ABC) transporter family ion channel.

5 Concluding Remarks

All atom DMD simulations with Medusa force field allows us to perform redesign protein structures with flexible backbone algorithm, which in turn permits evaluation of the impact of mutations on stabilities of proteins. Furthermore, coupling DMD with Medusa allows one to improve the homology models and refine the structures.

References

1. F. Ding, and N. V. Dokholyan, *Simple but predictive protein models*, Trends Biotechnol. **23**, 450–455, 2005.
2. N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, *Discrete molecular dynamics studies of the folding of a protein-like model*, Folding and Design **3**, 577, 1998.
3. F. Ding, and N. V. Dokholyan, *Emergence of protein fold families through rational design*, PLoS computational biology **2**, e85, 2006.
4. S. Yin, F. Ding, and N. V. Dokholyan, *Eris: An automated estimator of protein stability*, Nature methods **4**, 466–467, 2007.
5. S. Yin, F. Ding, and N. V. Dokholyan, *Modeling backbone flexibility improves protein stability estimation*, Structure **15**, 1567–1576, 2007.
6. J. Hao, A. W. R. Serohijos, G. Newton, G. Tassone, Z. Wang, D. C. Sgroi, N. V. Dokholyan, and J. P. Basilion, *Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers*, Public Library of Science Computational Biology, in press (2008).

7. A. W. R. Serohijos. T. Hegedus, A. A. Aleksandrov, L. He, L. Cui, N. V. Dokholyan, and J. R. Riordan, *Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function*, Proceedings of the National Academy of Sciences of the United States of America **105**, 3256–3261, 2008.

Predicting Protein Interactions from Functional Specificity

K. Anton Feenstra[1], **Giacomo Bastianelli**[2], **and Jaap Heringa**[1]

[1] Centre for Integrative Bioinformatics VU (IBIVU), Vrije Universiteit Amsterdam,
De Boelelaan 1081A, 1081 HV Amsterdam, The Netherlands
E-mail: {feenstra,heringa}@few.vu.nl

[2] Institut Pasteur, 25-28 rue du Dr Roux, Paris, France
E-mail: gbastian@pasteur.fr

Many protein families involved in protein-protein interaction (PPI) contain sub-families that interact with different protein binding partners. We have applied the Sequence Harmony method (*SH*), which is able to detect specificity sites from a multiple sequence alignment (MSA) containing sub-families. The input was a dataset of MSAs of interacting protein families, each time containing a set of non-interacting paralogous sequences. Exploiting the differences in sequence conservation between the binding and non-binding groups by *SH*, we demonstrate that predicted specificity residues turn out to reside on the protein surface. We also show that we can select interface residues with approximately 14% coverage (true-positive rate) at 27% error (false-positive rate).

1 Introduction

Specificity is a critical ingredient in regulation and signaling processes in cellular systems and it is most often achieved by recognition between specific proteins. Detection of specificity residues is most often used to pinpoint functional residues in general,[1] however, if the functional difference is based on protein-protein interactions (PPI), specificity sites can actually correspond to the interface region.

We will attempt to identify PPI interface regions from protein sequence using the previously introduced *SH* method for detection of subtype specific sites.[2,3] The general scheme of our approach is shown in Fig. 1, and is based on the availability of data on interacting (*A*-*B*) and non-interacting (*A*-*B*′) paralogs. The first step is the detection of a non-interacting ortholog (*B*′) to one of the interacting proteins (*B*, in this case). Second is the addition of orthologs to both the interacting and non-interacting proteins, and finally the selection of specificity residues using this pair of paralogous groups of orthologs.

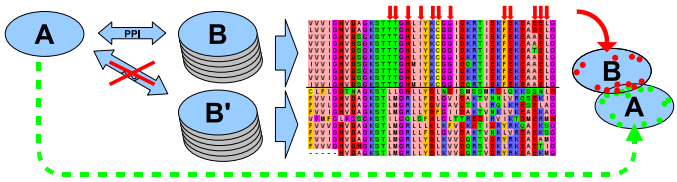


Figure 1. For a pair of interacting proteins (*A*-*B*) we find a non-interacting homolog (*B*′, no interaction with *A*), and orthologs of the interactor (*B*) and non-interactor (*B*′). Specificity residues for these ortholog groups (*B vs. B*′) are putative interface residues in *B*. Analogously, we select putative interface residues for *A* (using *A*′ that does not interact with *B*), which can be matched up to form the interface between *A* and *B*.

2 Methods

2.1 Homolog Detection

Experimental PPI evidence is based on the sum of socio-affinity scores from Gavin et al.[4] and Krogan et al.[5], as proposed by Van Noort et al.[6]. We took a score above a threshold of 4 as interacting, and below as non-interacting.

For paralog detection, we `blast` each interacting protein (*B*) against the `nr` database with `entrez_query=Saccharomyces cerevisiae`. We find the first *non-interacting* protein (B'), ordered according to the blast score (most similar first). We filter the hits by sequence length of least 80% of the query. For the ortholog search, we use the same rationale as for the paralog search but now using the Fungi database, and retain one hit for each of the first 10 organisms (so, at most 10 different organisms). For these hits, we require the length to be between 80% and 140% of the interacting protein, a bit-score of > 50 and an *e*-value of $< 10^{-15}$.

2.2 Protein-Protein Interaction Specificity

The basis of the interaction specificity residue detection is the selection of specificity residues using *SH*. These selections are further filtered using simple rules based on group and combined entropies.

For the *SH* method, sequences are taken from an input alignment and separated into user-specified groups. The *SH* score for two groups *A* and *B* is calculated as $SH = \frac{1}{2}\left(S_{A+B} - S_A - S_B\right)$, using the group entropies ($S = -\sum p \log p$) and combined entropy over the column ($S_{A+B} = -\sum \left(p_A + p_B\right) \log \left(p_A + p_B\right)$), where the sums are over all residue types. *SH* values range from zero for completely non-overlapping residue compositions, to one for identical compositions. Further details on the method were described previously.[2,3,7]

Our dataset contains 15 heterodimer complexes of *S. cerev.* proteins of size 30 or larger for which PDB files are available, and the analysis includes all $15 \cdot 2$ chains. Interface and surface residues were identified using `MSMS`.[8]

Interface	Surface	Criterion	Description
69.6%	50.2%	$S_B < S_{B'}$	more conserved in interactor
9.8%	30.7%	$S_{B'} < S_B$	less conserved in interactor
10.1%	10.2%	$S_{A+B} = 0$	conserved
34.3%	19.2%	$S_{B'} - S_B > 0.3$	much more conserved in interactor
9.5%	7.4%	$S_B = 0$ and $S_{B'} = 0$ and $SH = 0$	conserved in group but not between
48.9%	52.1%	$S_{A+B} > 0.4$	variable (non-conserved)
61.8%	67.0%	$S_{A+B} > 0.4$ within $S_{B'} - S_B > 0.3$ and $S_B < 0.3$	variable within conserved interactor
21.2%	12.9%	$S_{A+B} > 0.4$ and $S_{B'} - S_B > 0.3$ and $S_B < 0.3$	variable and conserved interactor
42.3%	33.8%	$SH < 0.2$	specific (low harmony)
28.9%	17.9%	$S_B < S_{B'}$ and $SH < 0.2$	conserved interactor and specific
17.8%	13.4%	$S_{B'} - S_B > 0.3$ and $SH < 0.2$	more conserved interactor and specific

Table 1. Behaviour of residues at the interface and at the rest of the surface (non-interacting part) in terms of *SH* and group (S_A, S_B) and overall (S_{A+B}) entropies.

3 Results and Discussion

We have observed (data not shown) that our predictions are very sensitive to a propor selection of paralog and ortholog sequences. Particularly, the prediction quality depends on the distance between the interacting protein and its non-interacting paralog.

3.1 (Non-)Interacting Paralogs

Our selection procedure was set up to identify non-interacting paralogs based on specific `blast`-searches. We observed, however, that less than 10 percent of all paralogs found were actually interacting. This opens the interesting possibility of applying our approach without the support of (high-quality) data on *non-interaction*, which in general is much less reliable than *interaction* data. This could be done on a much larger scale from available genomic sequence data and high-throughput PPI screens.

3.2 Surface and Interface Properties

Previously, we had already observed that specificity residues are predominantly located at the protein surface[2,3]. Since interface regions obviously are on the surface as well, we therefore analyzed the selection of surface residues in the current dataset. We used $SH \leq 0.3$, $S_B \leq 0.3$, and $S_{B'} - S_B \geq 0.3$ and find that between 71% and 95%, on average $86\% \pm 8$, are at the surface.

From the properties listed in Table 1, we can make a detailed comparison between interface residues and other surface residues (non-interface). We see an enrichment on the interface of residues with entropy S_B (interacting) lower than $S_{B'}$ (non-interacting), which is 70% over 50%. This ratio becomes bigger if the entropy difference is bigger, 34% over 19%. Also low-harmony sites are somewhat more prevalent on the interface. Conservation and variability seem to be equally distributed between the intface and rest of the surface.

3.3 Selecting Interface Residues

From the observed statstics, we will now derive some rules for the selection of interface residues. First of all, the entropy S_B (interacting) should be lower than $S_{B'}$ (non-

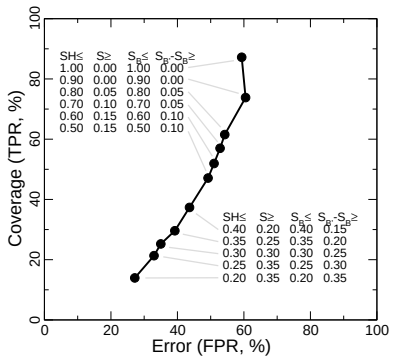


Figure 2. ROC curve (TPR *vs*. FPR) for selection of interface residues from different selection criteria.

interacting). This is true for 70% of interface residues, and for 34% the difference is even bigger (> 0.3). In addition, the *SH* score should not be high for a residue that is critical for interaction with the partner, but it also is not necessarily very low (42% with $SH < 0.2$).

In Fig. 2 we show a ROC plot for a set of different combinations of selection parameters, *cf.* Table 1, going from relaxed to more stringent. We can obtain very high coverage (90%), but at relatively high error (60%). Lowering the error rate goes sharply at the cost of coverage, and the lower half of the ROC curve even goes below the line expected for a random selection (where FPR $\sim$ TPR).

4 Conclusion

We have shown that subtype specificity can be used as a tool to pinpoint interface residues. However, care should be taken in selecting the paralog and ortholog proteins. Nevertheless, the trends we observed will be helpful in the further development of protein interface prediction. In addition, our analysis yields some insights in possible evolutionary selection mechanisms that have helped shape protein interfaces, and furthers our understanding of specificity of protein-protein interactions.

Acknowledgments

We thank Vera van Noort and Martijn Huynen for the Gavin-Krogan socio-affinities dataset. We thank Bas Dutilh and Martijn Huynen for stimulating discussions.

References

1. J. C. Whisstock and A. M. Lesk, *Prediction of protein function from protein sequence and structure*, Quart. Rev. Biophys., **36**, no. 3, 307–340, 2003.
2. K.A. Feenstra, W. Pirovano, K. Krab, and J. Heringa, *Sequence harmony: detecting functional specificity from alignments*, Nucleic Acids Res., **35**, no. web server issue, W495–W498, 2007, `http://www.ibi.vu.nl/programs/seqharmwww`.
3. W. Pirovano, K. A. Feenstra, and J. Heringa, *Sequence Comparison by Sequence Harmony Identifies Subtype Specific Functional Sites*, Nucl. Acids Res., **34**, no. 22, 6540–6548, 2006.
4. A.C. Gavin, P. Aloy et al., *Proteome survey reveals modularity of the yeast cell machinery*, Nature, **440**, 631–636, 2006.
5. N.J. Krogan, G. Cagney et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae*, Nature, **440**, 637–643, 2006.
6. V. van Noort, B. Snel, and M.A. Huynen, *Exploration of the omics evidence landscape: adding qualitative labels to predicted protein-protein interactions.*, Genome Biol., **8**, R197, 2007.
7. K. A. Feenstra, W. Pirovano, and J. Heringa, *Sub-type Specific Sites for SMAD Receptor Binding Identified by Sequence Comparison using "Sequence Harmony"*., in: From Comput. Biophys. to Syst. Biol., U.H.E. Hansmann, J. Meinke, S. Mohanty, and O. Zimmermann, (Eds.), vol. 34 of *NIC Series*, pp. 73–78, John von Neumann Institute for Computing, Jülich, 2006.
8. M.F. Sanner, A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces.*, Biopolymers, **38**, 305–320, 1996.

Short Membrane Proteins from Viruses: Channel-Pore Dualism?

George Patargias[1], Hugo Martay[1], Jens Krüger[2], Chin-Pei Chen[2], Yi-Ting Wang[2], Ming-Da Lee[2], Cheng-Chang Chen[2], Guo-Sheng Hong[2], and Wolfgang B. Fischer[2]

[1] Department of Biochemistry, University of Oxford, Oxford OX2 3QU, UK

[2] Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang-Ming Univeristy, Taipei 112, Taiwan
E-mail: wfischer@ym.edu.tw

The genomes of some viruses encode small membrane proteins which are known to alter membrane permeability by forming ion conducting pores. The alteration of the electrochemical gradient as a consequence of 'channel activity' has large scale consequences as it induces the fusion and budding process of the virion. Viral channel or pore forming proteins exist with different numbers of transmembrane (TM) domains. The working hypothesis is that the proteins diffuse as monomers in the lipid membrane and finally self-assemble to form the functional unit. Self-assembly has to take place at the level of the tertiary and quaternary structures within the low dielectric medium of the lipid membrane. Computational techniques are used to analyze the assembly process. Ion flux is simulated using steered molecular dynamics (MD) simulations and analyzed using Langevin equation of motion. Conductance measurements flank the in silico investigations. Data of the channel forming protein Vpu from HIV-1 will be shown as a test case.

1 Introduction

In a series of viruses short membrane proteins have been identified which are found to alter permeability of host cell membranes[1]. For some of them such as M2 from Influenza A, functional analysis have been done and it is now established that M2 acts as a proton channel. For others proteins such as Vpu from HIV-1, p7 from HCV, and others data on ion conductance were obtained by reconstituting the proteins into artificial bilayers. In analogy to M2 however it is concluded that these proteins act as channel forming proteins in vitro as well. For Vpu from HIV-1 for example increasing evidence is now given in the literature that the protein interacts with host cell factors and hampers their mode of action. In other words besides altering the permeability of the lipid membrane Vpu acts also as a cellular modulator via assembling with host cell proteins. No matter which role the proteins accomplish they have to assemble to fulfill their role in the life cycle of the virus either with themselves to form channels or pores or with host cell factors. In the light of these findings on a molecular level the question arises (i) once assembled, how selective are these proteins, and (ii) what is the mechanism of assembling per se. These questions are also relevant for other diffusion processes across the lipid membrane such as viral fusion, DNA/RNA delivery and even budding. Structural information of these proteins is emerging via NMR spectroscopic investigations. For Vpu they have confirmed that the TM domain is helical. Computational methods may serve as a torch to enlighten atomic details, since high resolution data is not available for all of the proteins.

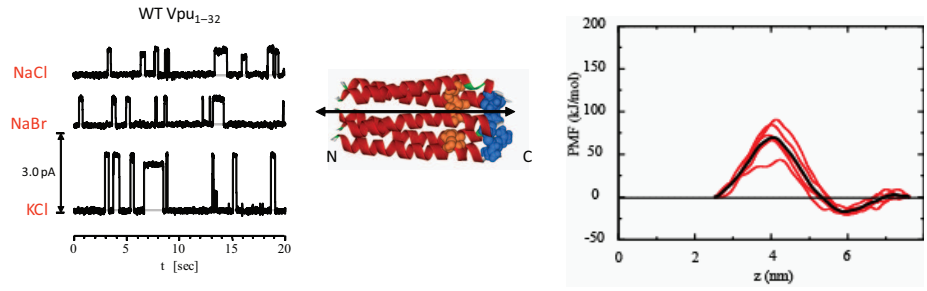


Figure 1. Bilayer recordings of Vpu_{1-32} using different types of salt at a holding potential of 100 mV, 500 mM salt concentration, HEPES, pH 7, reconstituted into POPE : DOPC (1:4) (left). Pentameric model of Vpu_{1-32} (middle) through which the ions are pulled. Respective potential of mean force of a K-ion being pulled through the pore. In black the averaged values for 5 (red) individual pullings.

2 Results

2.1 Experiments

Experimental findings with a peptide containing the first 32 amino acids of Vpu, Vpu_{1-32} (MQPIPIVAIV$_{10}$ ALVVAIIIAI$_{20}$ VVWSIVIIEY$_{30}$ RKI), reconstituted into artificial lipid bilayers show conductance in the range of 17 - 20 pS in 300 - 500 mM KCl solution[2]. When changing the anion, e.g. NaCl to NaBr a small change in conductance is found compared to a larger change when changing from NaCl to KCl. This indicates that the assembled bundle prefers to conduct cations over anions. However in the light of conductance measurements with a series of other monovalent chloride salts it is concluded that the bundle is only a weakly selective channel. Together with the findings of altered substrate permeability by other groups it is concluded that the bundle assembly may be able to conduct ions AND small molecules. The mechanism of how the bundle discriminates is still not clear. It is suggested, that conformational changes due to lipid composition are responsible.

2.2 Computer Simulations

To transfer the experimental data into in silico experiments helical models of a pentameric assembly of Vpu_{1-32} have been generated using XPLOR (A. Bruenger). Using a helical motif for the TM domain is suitable based on NMR experiments. The bundle model has a hydrophilic C terminus due to Ser-24 and Arg-31 facing the lumen of the pore and a hydrophobic stretch towards the N terminus. With steered molecular dynamics simulations in a fully hydrated lipid bilayer using Gromacs 3.2.1 (Gromos 96 ff, PME) K-, Na- and Cl-ions have been pulled through the lumen of the pore. Assuming a model of one-dimensional stochastic dynamics for the ion Langevin equation is used to calculate the potential of mean force of the ion. The data indicate a s-shaped curve with a minimum at the C terminus and a maximum within the pore in the region of the hydrophobic stretch produced by the amino acids towards the N terminus. A comparison of the energy bariers of the different ions shows a preference for cations over the anion. Based on the current

model and physical method to calculate the PMF the in silico data support the experimental data of Vpu being a weak channel.

3 Discussion

The quality of the current data depends on the physical model used for ion diffusion, the method used to derive the relevant data and the bundle model itself. To elucidate the affect of the bundle emphasis is currently given to establish a protocol for generating reliable bundle models. The approach can be summarized as following: (i) detection of the TM domain(s) of the protein using secondary structure prediction programs, (ii) obtaining an equilibrated structural model of the momoner[3], and (iii) generating high quality bundle models by screening conformational space. In this last stage a protocol is used which combines backbone positioning followed by side chain generation (Krüger and Fischer in this Proceedings p. 269). All the calculations are based on the assumption that the protein is produced as a monomeric unit in the endoplasmic reticulum and will fold and diffuse as a monomer prior to assembly.

4 Concluding Remarks

Self-assembly of the Vpu protein leads to a weak selective bundle, which may be also true for other viral proteins with a single TM domain. Depending on the experimental or in vivo conditions Vpu may also allow substrates to pass (channel-pore dualism). Correlation of in silico models with experimental data can be achieved to elaborate on the structure-function paradigm. Calculating functional data will be used to suggest the most reliable models.

Acknowledgments

WBF thanks the Government of Taiwan, NSC, NYMU, DAAD, Bionanotechnolgy IRC (Oxford, UK) for financial support. We also thank Paderborn Center for Parallel Computing PC_2 for providing computer time.

References

1. W. B. Fischer, *Protein Reviews: Viral membrane proteins: structure, function and drug design. Volume 1* (Kluwer Academic / Plenum Publisher, New York, NY, 2005).
2. T. Mehnert, A. Routh, P. J. Judge, Y. H. Lam, D. Fischer, A. Watts and W. B. Fischer, *Biophysical characterisation of Vpu from HIV-1 suggests a channel-pore dualism.*, Proteins **70**, 1488-1497, 2008.
3. J. Krüger, and W. B. Fischer, *Exploring the conformational space of Vpu from HIV-1: a versatile and adaptable protein.*, J. Comp. Chem., in print, 2008.

Coarse-Grained Molecular Models for High-Throughput and Multi-Scale Functional Investigations

Kay Hamacher

Technische Universität Darmstadt, Fachbereich Biologie, Institut für Mikrobiologie und Genetik, Schnittspahnstr. 10, 64287 Darmstadt, Germany
`http://www.kay-hamacher.de`
E-mail: hamacher@bio.tu-darmstadt.de

We discuss the development of a physics-based, coarse-grained molecular model and its subsequent application to in silico functional investigation. Such models overcome shortcomings of both sequence-based bioinformatics (no physics) and molecular dynamics simulations (large CPU requirements, therefore sampling issues) as an integrative approach. We demonstrate this claim by showing: capabilities to investigate macromolecular thermodynamics [here: assembly of the bacterial ribosome[1]] integrating information into the physical realm [here HIV protease[2] and drug resistance[3]] We discuss also further (technical) advantages: the computations are highly parallelizable, allow for high-throughput screening, and can be integrated with an information driven analysis technique[4].

1 Introduction

Understanding protein-protein interactions, signaling pathways, protein dynamics, and biomolecular mechanics in general is of uttermost importance in systems biology and in all areas of life science.. Proof of the great use of mathematical modeling of signaling-networks are well known, e.g. the work by Lee et.al. on the Wnt signaling for oncogenesis[5].

To this end established protocols rely on off-the-shelf bioinformatics algorithms such as Hidden Markov Models (HMM) and other machine learning approaches to investigate e.g. the vertices in interaction networks. It was, however, noted for example by Albert[6] that most computationally derived protein-protein-interactions networks show too many false positives. This can be attributed to several drawbacks of statistical modelling as the underlying theory of sequence-based bioinformatic procedures such as HMMs: these bioinformatics models are 'local' in the sense that they not necessarily take into account multi-amino-acids correlations - a property that contributes to recognition and structural stability as recently re-confirmed by the Plotkin lab[7].

Molecular dynamics[8] on the other hand allows for the detailed investigation of physical and chemical properties. The high computational costs are the major drawback of this approach: it is – despite exponential progress in hardware and software – still impossible to simulate large number of protein mutants, although such databases are routinely investigated by sequence-based algorithms.

Now reduced molecular models allow to bridge the gap between pure sequence-focused bioinformatics and the detailed molecular biophysics – allowing for real *in silicio* structural & functional proteomics and derivation of the above mentioned interaction networks due to the fast computation of reduced (thermo-)dynamics in those models. These models take into account physical interactions and their effects such as binding stability or responses

to perturbations of the protein's structure. Both the elimination of false positives and the inclusion of potential, not yet detected interactions will help to derive more precise protein interaction networks, perform myriads of thought experiments, and allow therefore the functional characterisation of the underlying genomes.

2 Reduced Molecular Models

Interactions within biomolecules are usually modelled with sohphisticated force-fields and contain various terms such as Lennard-Jones-potentials, electrostatics, hydrogen-bonding etc. For small perturbations such a full potential can be expanded to second order in a Taylor expansion, which then poses the problem of diagonalizing a quadratic form[9]. Eigenmovements and eigenfrequencies are then easily determined. Tirion[10] was the first to investigate the resulting dynamics. Extensions with respect to anisotropy[11], extended thermodynamics[12], or non-linear conformational transitions[13] were introduced in recent years.

3 Applications

3.1 Protein Aggregation Network in the Ribosome

Using Micheletti's Self-Consistent-Pair-Contact-Potential iteration scheme[12] we investigated the influence of the presence or absence of ribosomal proteins and pairs of them on the binding affinities of the others[1]. This is a protocol suitable for general biomolecular formation processes. For the organism *T. thermophilus* we found that the assembly map is in very good agreement with the experimental known one for *E. coli*. This study examplifies: reduced molecular models can be used in multiscale-investigations and enable one to derive computationally protein-protein-interaction networks.

3.2 Amino-Acid-Mutation Network in the HIV1-Protease

We were furthermore able to parameterize reduced models to incorporate effects of mutated amino acids[2]. Using this system we investigated the origin of the drug resistance mechanism of the V82F-I84V-mutant of HIV1-protease (induced increase of flexibility in protein-substructures). By computing "hybrid" mutants with systemtically changed wild-type-like and mutant-like interactions one can reveal the mechanism: this particular mutation weakens the interaction at the "joints" of the protease-dimer. With this we have shown for the first time that a *functional* annotation of mutation events is possible. Continuing this work we applied the method to some 40,000 mutants[3] and correlated the changes in the protein mechanics to the stability in sequence space. The dimerization interface in the protease turned out to be an interesting target for new protease inhibitors: it has a large influence on the efficient molecular dynamics, while at the same time there are evolutionary barriers so that drug resistance evolution is repressed.

4 Concluding Remarks

We have motivated a new approach to cope with shortcomings of well-established and widely known algorithms in computational biology: sequence-based bioinformatics (no insight into the molecular biophysics) and molecular dynamics (prohibitive high computational costs). With such reduced molecular models it is possible to merge the information space of genomes with the physical realm of proteomes - thus allowing for physical underpinning of molecular systems biology.

Two examples illustrate the usefullness of this approach: the assembly network of proteins in the ribosome and the amino-acid network during evolution of drug resistance.

Acknowledgments

We are indebted to the organizers of the CBSB08 for the opportunity to present our work. The underlying studies were financially supported by the Fonds der chemischen Industrie through a Liebig-Fellowship and through ongoing financial support.

References

1. K. Hamacher, J. Trylska, and J.A. McCammon, *Dependency Map of Proteins in the Small Ribosomal Subunit*, PLoS Computational Biology, **2**, e10, 2006.
2. K. Hamacher and J. A. McCammon, *Computing the Amino Acid Specificity of Fluctuations in Biomolecular Systems*, J. Chem. Theory Comput., **2**, 873–878, 2006.
3. K. Hamacher, Relating sequence evolution of hiv1-protease to its underlying molecular mechanics, 2008, Gene, accepted.
4. K. Hamacher, *Information Theoretical Measures to Analyze Trajectories in Rational Molecular Design*, J. Comp. Chem., **28**, 2576–2580, 2007.
5. E. Lee, A. Salic, et al. *The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway*, PLoS Biology, **1**, e10, 2003.
6. R. Albert, *Scale-free networks in cell biology*, J Cell Sci, **118**, 4947–4957, 2005.
7. M. R. Ejtehadi, S. P. Avall, and S. S. Plotkin, *Three-body interactions improve the prediction of rate and mechanism in protein folding models*, Proc. Nat. Acad. Sci., **101**, 15088–15093, 2004.
8. T. Schlick, *Molecular Modeling and Simulation*, Interdisciplinary Applied Mathematics - Mathematical Biology. Springer, New York, 2002.
9. W.H. Press et al, *Numerical Recipies in C*, Cambridge University Press, Cambridge, 1995.
10. M.M. Tirion, *Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis*, Phys. Rev. Lett., **77**, 1905–1908, 1996.
11. A.R. Atilgan, S.R. Durrell, R.L. Jernigan, et al. *Anisotropy of fluctuation dynamics of proteins with an elastic network model*, Biophys. J., **80**, 505–515, 2001.
12. C. Micheletti, J.R. Banavar, and A. Maritan, *Conformations of Proteins in Equilibrium*, Physical Review Letters, **87**, 088102–1, 2001.
13. O. Miyashita, J. N. Onuchic, and P. G. Wolynes, *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins*, PNAS, **100**, no. 22, 12570–12575, 2003.

Simulation of RNA Folding on the Simple Cubic Lattice

Shura Hayryan[1], **Karen G. Sargsyan**[2], **and Chin-Kun Hu**[1,3]

[1] Institute of Physics, Academia Sinica, Taipei 11529, Taiwan
E-mail: {*shura, huck*}*@phys.sinica.edu.tw*

[2] Yerevan Physics Institute, Alikhanian Brothers 2, 375036 Yerevan, Armenia

[3] Center for Nonlinear and Complex Systems and Department of Physics,
Chung Yuan Christian University, Chungli 32023, Taiwan

We consider the RNA molecule as a self-avoiding walk on the simple cubic lattice and use a modified Wang-Landau method to calculate its specific heat C_v and mean square end to end distance R^2 as a function of temperature. The energy function includes hydrogen bond energy ε, stacking energy $\overline{\varepsilon}$, or chain rigidity energy J. We find that inclusion of pseudoknots or $\overline{\varepsilon}$ greatly changes the behavior of C_v and inclusion of stacking energy greatly changes the behavior of R^2.

1 Introduction

RNA molecule is a highly charged heteropolymer built from 4 types of nucleotide bases. Like the proteins, a RNA molecule can attain secondary and tertiary structure with the latter being function-related. It is widely believed that the folding of a RNA molecule occurs in two stages. First the elements of the secondary structure are formed, and then such elements are folded to form the tertiary structure with the energy change much smaller than that in the first stage. This mechanism is known as hierarchical folding[1]. The typical physical interactions in RNA molecules include complementary hydrogen bonds between A-U and G-C pairs of nucleotide bases, strong electrostatic interactions and stacking interaction between neighbor bounded pairs[2,3]. The important feature of hydrogen bonding between bases is that this bonding is of saturated nature meaning that once two bases are bonded, there can be no other bonds with other bases.

Most theoretical models for predicting the secondary structures of RNA sequences neglect the formation pseudoknots (see e.g. Fig. 1 in Ref. 3) and assume the hierarchical mechanism of folding. Among other theoretical approaches, the coarse grain lattice models have been applied to study the statistical behavior of RNA folding, *e.g.* Leoni and Vanderzande[4] applied so called self avoiding two-tolerant trail model to study the RNA folding on a square lattice, in which the saturated nature of the hydrogen bonds is taken into account by allowing the chain to visit each bond at most twice. The twice visited bond corresponds to the H-bonded bases. They consider the simple Hamiltonian which includes only the energy of hydrogen bonds. With this model they obtained the phase diagram of the system which includes native, coiled and branched polymer phases. The similar two-tolerant trail model is applied for simulations on face centered cubic lattices[5] and Husimi lattice[6].

In this paper we model a RNA molecule as a self avoiding walk on the simple cubic lattice and use Monte Carlo method to study its thermodynamic behavior.

2 The RNA Model and Monte Carlo Moves

We consider a RNA molecule of N bases as a self-avoiding random walk of length N on a $3D$ simple cubic lattice. Each lattice site corresponds to one nucleotide base and can be visited only once. The lattice bonds mimic chemical links between monomers. We consider following interactions: (i) *Hydrogen bonds between two bases*. The homopolymeric approximation is assumed wherein each base can be H-bonded to any other if their distance equals to 1 lattice constant. This approximation is not principal, and the procedure described below can be easily extended to the heteropolymeric case. However, here the saturated nature of H-bonding is maintained meaning that each base is allowed to be bonded only with one other. The energy value ε is associated with each bounded base-pair. Due to the homopolymeric approximation, here the value of ε is taken as an average of the energies of A-U and G-C bonds. We use the value[2] $\varepsilon = 15kcal/mol$. (ii) *Stacking interactions between two successively H-bonded pairs*. When two neighbor base-pairs are H-bonded then an energy $\overline{\varepsilon}$ is added to the Hamiltonian. We take[2] $\overline{\varepsilon} = 0.5\varepsilon$ or ε. (iii) *Bending energy or the rigidity of the chain*. Bending of the chain is penalized by some energy J. Some unit vectors $\overrightarrow{s}_i$ are assigned to each lattice bond to describe this term. With these basic interactions included, we have the Hamiltonian

$$H = -\sum_{(i,j)\in S} \varepsilon - \sum_{(i,j)\in S} \overline{\varepsilon} g_{i,i+1,j-1,j} - J\sum_{i=1}^{N-1} \overrightarrow{s}_i \overrightarrow{s}_{i+1}. \tag{1}$$

Here the three sums correspond to the three types of the interactions describe above. In the first and the second terms the summation goes over all base pairs (i,j). The function $g_{i,i+1,j-1,j}$ in the stacking energy term takes value 1 if the base pairs (i,j) and $(i+1,j-1)$ are H-bonded simultaneously, and 0, otherwise. The sum in the third term goes over all chemical bonds. At this point we purposely don't include the electrostatic interactions which will be discussed later.

To investigate the thermodynamics of our system we use the Monte Carlo technique to sample the phase space. Starting from some initial configuration of the chain we apply the pivot algorithm[7] as a proposal move to the next configuration. In the initial configuration the hydrogen bonds are assigned starting from some random site first in one direction, and then in the opposite one. The basic procedure is as this: if the chosen site is not H-bonded then a random search is made between its nearest neighbors (not chemically connected to the given site). When the first non-bonded site is found then a bond is placed between it and the chosen one. If no free neighbors are found then the next site is considered. Two strategies are applied for assigning the hydrogen bonds during the Monte Carlo move: (i) When the center for pivot rotation is chosen the hydrogen bond of the given site as well as of the next and the previous sties are considered broken. After the pivot move these bonds are restored again using the procedure described above. After restoring the bonds the energy of the new configuration is calculated. (ii) All hydrogen bonds along the chain are considered broken and reconstructed anew after the pivot move. The Wang-Landau[8] algorithm is used for calculating the density of states. Then the average thermodynamic or the structural quantities are being calculated with the general formula

$$\langle A\rangle_T = \frac{\sum_i \langle A\rangle_{E_i} g(E_i) e^{-E_i/kT}}{\sum_i g(E_i) e^{-E_i/kT}}. \tag{2}$$

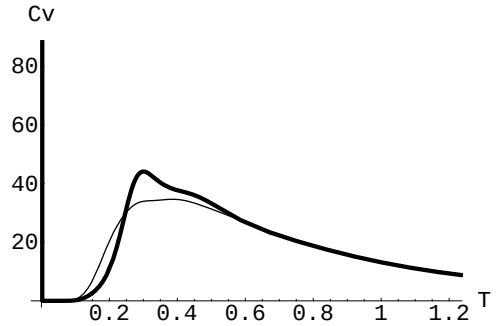


Figure 1. The specific heat C_v as a function of the temperature T (in unit ε/k) for the system with only hydrogen bond energy and including pseudoknots (bold line) or excluding pseudoknots (light line), $N = 100$.

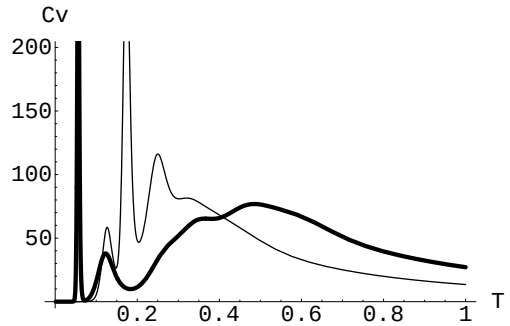


Figure 2. C_v as a function of T (in unit ε/k) for the system with hydrogen bonding & stacking interactions of $\overline{\varepsilon} = 0.5\varepsilon$ (bold line) or $\overline{\varepsilon} = \varepsilon$ (light line), and without pseudoknots, $N = 100$.

Here T is the temperature, k is the Boltzmann constant, $\langle A \rangle_T$ is the quantity under consideration, E_i is the energy of the state i, and $g(E_i)$ is the density of state. In the simulations, we allow formation of the pseudoknots[3].

3 Results

Before starting simulations with the full Hamiltonian (1), we have studied the behavior of the systems with simplified energy functions in order to get more detailed insight. Due to the limited space we bring here the results for only these simplified systems.

The system with the hydrogen bond energy only. This corresponds to the first sum in (1). The simulated results of specific heat C_v as a function of the temperature T (in unit ε/k) for the system with pseudoknots (bold line) or without pseudoknots (light line) are plotted in Fig. 1, which shows a well defined compaction transition; the transition is sharper in the system with pseudoknots. The same transition is seen also on the graph of the mean squared end-to-end distance (not shown here).

Hydrogen bonds plus stacking energy. This corresponds to the sum of the first two terms in (1). In this case (Fig. 2) the specific heat is not single-peaked. While the small

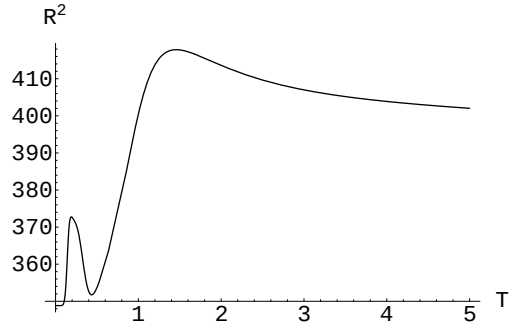


Figure 3. The mean square end to end distance R^2 vs. T (in unit J/k) with $\overline{\varepsilon} = 0$, $J = 2kJ/mol$ and $N = 100$.

peak at the left side of the main one might be non-essential (within the sampling error) the one at the right side can not be overlooked; it might be the manifestation of two-step RNA folding in the hierarchical folding mechanism. Figure 2 shows that changing the stacking energy affects considerably the position of the peak.

Hydrogen bonds plus chain rigidity. This corresponds to the sum of the first and third terms in (1), where J may be estimated by investigating the behavior of the correlation length for the Hamiltonian at room temperature. Some estimates show that the correlation length is not larger than several thousands of monomers which gives the value between $0.5kJ/mol$ and $4kJ/mol$. For this case we plot the mean squared end-to-end distance as a function of T in Fig. 3 with $J = 2kJ/mol$. One can see that even such a small rigidity brings a strong frustration into the system; the small peak on the graph manifests the existence of glassy structure at low temperatures.

Acknowledgments

This work was supported by Grants NSC 96-2911-M 001-003-MY3 & AS-95-TP-A07, and National Center for Theoretical Sciences in Taiwan.

References

1. I. Tinoco and C. Bustamante, J Mol Biol **263**, 271, (1999).
2. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
3. Y. S. Mamasakhlisov, S. Hayryan, V. F. Morozov, and C.-K. Hu, Phys. Rev. E **75**, 061907 (2007).
4. P. Leoni and C. Vanderzande, Phys Rev E **68**, 051904 (2003).
5. M. Baiesi, E. Orlandini and A. L. Stella, arXiv:cond-mat/0303226v1.
6. R. A. Zara and M. Pretti, Physica A **371**, 88 (2006).
7. N. Madras, G. Slade, *The self-avoiding walk* (Birkhäuser, Boston, 1993).
8. F. G. Wang and D. P. Landau, Phys Rev Lett **86**, 2050 (2001); Phys Rev E **64**, 056101 (2001).

Designing an Automatic Pipeline for Protein Structure Prediction

Sebastian Kmiecik[1], Michal Jamroz[1], Anna Zwolinska[1], Pawel Gniewek[1], and Andrzej Kolinski[2]

[1] Selvita Sp. z o.o.,
Ostatnia 1c, 31-444 Cracow, Poland
E-mail: sebastian.kmiecik@selvita.com

[2] Laboratory of Theory of Biopolymers,
Faculty of Chemistry, University of Warsaw,
02-093 Warsaw, Poland

Building accurate 3D structural models of proteins and protein assemblies is a challenging task. Our modeling technology is based on the CABS model, extensively tested, state-of-the-art approach to protein structure prediction. The modeling process is divided into two stages: CABS fold assembly followed by the model refinement/selection procedure, using an all-atom representation and a more exact interaction scheme enabling high resolution structure prediction. Fold assembly can be done in a framework of a standard comparative modeling procedure, where spatial restraints are derived from alternative sequence alignments with a template/templates. Preferentially in more difficult modeling cases, a new approach to comparative modeling can be used, which does not require the prior alignment. Selvita's goal is to provide an integrated tool-kit for automated protein structure predictions. However, like blind prediction experiments show, due to high complexity of prediction tasks, fully automated approach often doesn't guarantee the highest possible performance. Therefore, human intervention is made possible at every stage of modeling.

1 Introduction

Thanks to international effort in the genome sequencing projects, enormous library of protein sequences is now available. Despite extensive efforts in structural genomics, the number of experimentally determined protein structures, typically by costly X-ray crystallography or NMR spectroscopy procedures, is lagging far behind the number of known protein sequences. Since proteins are involved in practically all functions performed by a cell, knowledge of protein structures is necessary for understanding and controlling molecular mechanisms of life. Current assumptions are, that for a large fraction of proteins whose structures will not be determined experimentally, computational methods can provide valuable information[1].

2 Multiscale Approach to Structure Prediction: Comparative Modeling and Fold Recognition

During computational protein structure determination the following main challenges can be identified: 1) High accuracy structure prediction, at the resolution comparable to experimental methods, to enable predicted models utilization in a number of protein structure-based approaches (e.g. drug design, protein design, molecular docking, molecular replacement), which is now possible in Comparative Modeling (CM) cases[2], 2) Structure prediction of proteins or protein fragments for which sequence search methods failed to find

unambiguous homologs with known structure (Fold Recognition (FR) and New Fold (NF) prediction)

To meet criteria of both challenges, precise interaction scheme, sensitive to small atomic rearrangement, should be somehow combined with high efficiency in exploring proteins conformational space. That can be achieved by combining all-atom and reduced modeling: the multiscale modeling. Properly designed reduced models make possible very effective search of the protein's conformational space[3] and all-atom modeling enable exact scoring and refinement of the models. Our modeling technology is based on a such hierarchical approach[2]. Reduced-space search of the conformational space by the CABS[3] is followed by a reliable transition into the all-atom resolution and by subsequent fine-tuning and assessment of the final models. Such multiscale approach enable high-resolution protein structure predictions, predictions of protein interactions[4], computer-aided drug design and even study of protein dynamics[5].

CABS computational technology has been rigorously tested during CASP6 (Critical Assessment of Techniques for Protein Structure Prediction) world-wide experiment by the Kolinski-Bujnicki group, which ranked second best among over 200 groups participating, and ranked first when the consistency of the prediction was used as a criterion (the number of CASP targets placed in the top 20 of the best predictions)[6].

The design of CABS model enable easy implementation of spatial restraints. Such restraints can be derived by a large number of bioinformatics tools from appropriate known structures or from experimental sources e.g. from sparse NMR data. Therefore, essentially the same approach is possible at various levels of protein modeling difficulty from CM, to FR and NF cases. For the sake of flexibility two basic modeling pathways were designed and one alternative to make the prediction more effective. The entire prediction pipeline could by briefly outlined as follows (see the flowchart in the Figure 1): 1) Pre-processing: Template identification, secondary structure prediction, target- template alignments, input for more sophisticated user defined FR multiple alignments, 2a) Fast modeling track (easy CM cases) including fast scoring of alternative alignments and generation of spatial restraints, 2b) Rigorous modeling track (hard CM and FR cases) including 3D threading and generation of spatial restraints, 2c) Alternative modeling track by TRACER (hard CM and FR cases) - without prior alignments[7], 3) CABS modeling, 4) Post-processing: trajectory clustering, selection of clusters representatives, rebuilding from reduced to all-atom representation and finally all-atom models refinement and ranking.

Additionally in the most difficult cases (NF) ab initio modeling based only on target sequence can be performed (the accuracy of the resulting models is sometimes sufficient for structure-based protein function identification).

3 Automatic or Human Driven?

As blind structure prediction experiments demonstrated, human expert experience and intuition becomes a key point to the best possible performance, especially in difficult CM and FR[1]. Also in high resolution structure prediction, when a fraction of an Angstrom of the final model resolution matters, human intervention may by helpful by manual insertions of a template structure fragments into the final model. However, our goal is to develop fully automated structure prediction protocol which enable structure prediction on a genomic scale. Considering difficult modeling cases, the modeling approach without prior align-

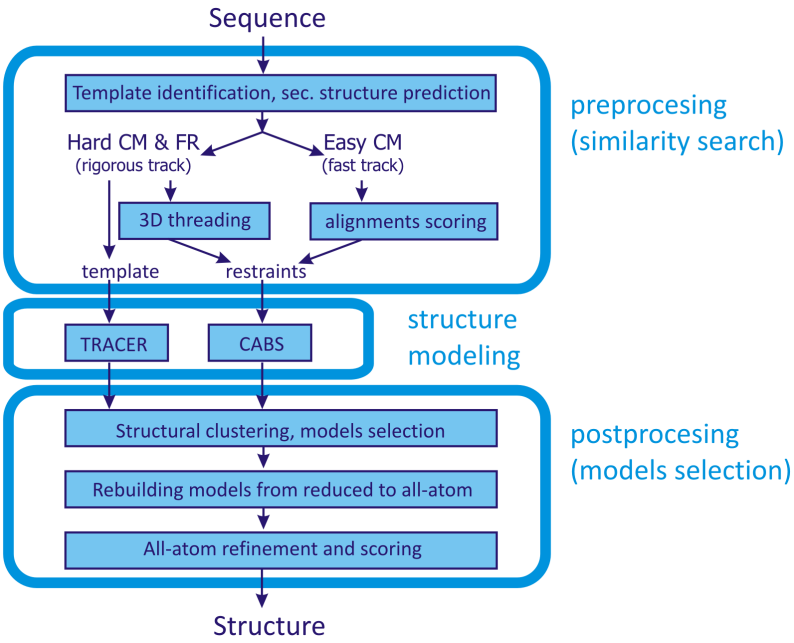


Figure 1. The protein structure prediction flowchart - see the text.

ments[7], included in our pipeline, seems to be an extremely promising step towards fully automated modeling (errors in alignments seem to be the main source of failures in protein structure prediction[1]).

References

1. O. Schueler-Furman, C. Wang, P. Bradley, K. Misura, D. Baker, *Progress in modeling of protein structures and interactions*, Science **310**, 638-42, 2005.
2. S. Kmiecik, D. Gront, A. Kolinski, *Towards high-resolution protein structure prediction. Fast refinement of reduced models with all-atom force field.*, BMC Structural Biology , 7:43, 2007.
3. A. Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochim. Pol. **51**, 349-371, 2004.
4. M. Kurcinski, A. Kolinski, *Hierarchical modeling of protein interactions*, J Mol Model **13**, 691-8, 2007.
5. DA Debe, JF Danzer, WA Goddard, A. Poleksic, *STRUCTFAST: Protein sequence remote homology detection and alignment using novel dynamic programming and profile-profile scoring*, Proteins **64**, 960-967, 2006.
6. S. Kmiecik, A. Kolinski, *Characterization of protein-folding pathways by reduced-space modeling.*, Proc Natl Acad Sci USA **104**, 12330-5, 2007.
7. A. Kolinski, D. Gront, *Comparative modeling without implicit sequence alignments.*, Bioinformatics **23**, 2522-27, 2007.

Conformational Polymorphism of a Fibrillogenic Fusion Peptide in Explicit Solvent and at an Interface

Volker Knecht and Reinhard Lipowsky

Max Planck Institute of Colloids and Interfaces, Research Campus Golm, 14424 Potsdam, Germany
E-mail: vknecht@mpikg.mpg.de

The development of therapeutic agents against amyloid diseases requires an understanding of the conformational properties of fibrillogenic species on a microscopic level. Here we have used molecular dynamics simulations to study the fibrillogenic fusion peptide B18 in monomeric form in various environments. In particular, our results indicate that B18 forms β-sheet/coil conformations in water, whereas adsorption at a water/vapor interface induces α-helical conformations. α-β transition pathways in water are suggested. Our results show that previous spectroscopic measurements reflect properties of monomers and reveal a pronounced conformational polymorphism of B18 in different environments, thus highlighting the need for an explicit description of the solvent environment in order to understand fibrillogenic species from molecular dynamics simulations.

1 Introduction

A number of neurodegenerative diseases such as Alzheimer's are associated with the conversion of proteins from a soluble, functional form into a β-rich structure that is highly prone to aggregate into toxic oligomers or so-called amyloid fibrils. The conformational transition is believed to take place in a partially denaturing environment of a cellular compartment either in solution or at an interface. Hence, the development of therapeutic agents against amyloid diseases requires an understanding of the conformational polymorphism in different environments on a microscopic level. To study fibrillogenic species experimentally in atomic detail is difficult because of their tendency to aggregate. Therefore, an indispensable tool to study these systems is provided by computer simulations. Here we have chosen the 18-residue peptide B18, a fragment of the sea-urchin fertilization protein Bindin[1] as a model system. The amphiphilic sequence of the peptide is shown in Fig. 1 (a). B18 forms amyloid fibrils *in vitro*. The soluble form in water forms β-sheet and coil structures with increased β-sheet content in the presence of NaCl as indicated from circular dichroism (CD) spectroscopy. Addition of trifluorethanol (TFE) or adsorption of B18 at a water/air interface induce α-helical conformations as indicated from CD or infrared reflection absorption spectroscopy (IRRAS), respectively, the latter suggesting helices to be parallel to the interface.[2,3] Nuclear magnetic resonance (NMR) measurements indicate a helix-kink-helix motif for B18 in water/TFE with 70:30 volume fractions[2] as shown in Fig. 1 (b). No detailed structure is available for B18 in pure water in the presence or absence of NaCl or at a water/air interface. In fact, it is not even clear to which extent conformations indicated from CD or IRRAS arise from mono- or oligomers. We have used molecular dynamics (MD) simulations to study B18 in monomeric form in explicit solvent and interfacial environments on a microscopic level.[4]

Figure 1. (a) Amino acid sequence and (b) structural model for B18 in a water/TFE 70:30 (volume fractions) mixture based on NMR data[2] in ribbon representation. The color code distinguishes between hydrophobic (*yellow*) and hydrophilic (*blue, green, and red*) residues.

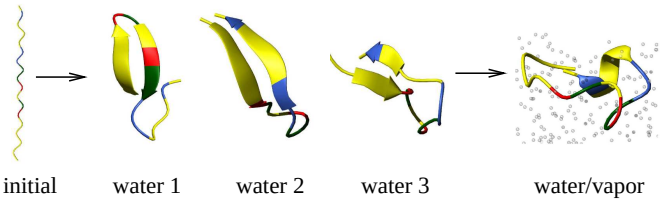


Figure 2. Initial configuration ("initial"), typical β-sheet conformations of B18 in water ("water $1-3$"), and representative conformation at a water/vapor interface ("water/vapor"). The representation is similar to that chosen in Fig. 1 (b). In "water/vapor", positions of water oxygens are indicated as gray dots.

2 Methods

B18 in explicit water/TFE with 70:30 volume fractions, pure water in absence and presence of 100 mM NaCl, and at a water/vapor interface with vapor modeling air were studied using periodic boundary conditions. Initial peptide configurations were a helix-kink-helix as in Fig. 1 (b), an extended as in Fig. 2 (*initial*) or a β-sheet/coil conformation as in Fig. 2 (*water 3*). Typically, three 50 ns simulations using the same initial peptide configuration, but different sets of initial velocities were performed. Most simulations were carried out at 293 K to mimic experimental conditions. Simulation details are given elsewhere.[4]

3 Results and Discussion

B18 in water/TFE with 70:30 volume fractions, pure water in the absence or presence of 100 mM NaCl, and at a water/vapor interface with vapor modeling air were studied. During simulations started from extended peptide configurations as in Fig. 2 (*initial*), coil and β-sheet conformations as in Fig. 2 (*water 1–3*) were adopted. β-sheets were mainly formed by hydrophobic residues, see Fig. 2 (*water 1–3*) (*yellow*). Addition of NaCl led to an increase in the average β-sheet content. Various β-sheets containing different residues were observed. β-sheets formed twice in independent simulations, suggesting that they are typical structures in water, are shown in Fig. 2 (*water 1–3*).

Pre-formed α-helical conformations as in Fig. 1 (b) were more stable in water/TFE or at a water/vapor interface than in pure water. In general, α-helical conformations were more stable in the C-terminal than the N-terminal half of the peptide. In water at 350 K, transitions from α-helical into β-sheet conformations where observed as shown in Fig. 3.

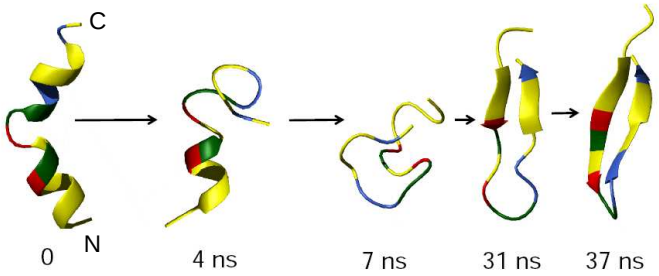


Figure 3. α-β transitions of B18 in water at 350 K. The representation is similar to that chosen in Fig. 1.

The C-terminal and, after 7 ns, the N-terminal helix dissolved in two independent simulations. Thereafter, the peptide underwent a transition from a compact coil as in Fig. 3 (*7 ns*) to more extended β-sheet conformations, see Fig. 3 (*31 – 37 ns*). The occurrence of a compact coil intermediate is similar to what has been observed during α-β transitions of other sequences in previous simulations, see references in[4], suggesting a universal feature for α-β transitions.

A peptide in β-sheet/coil conformation as in Fig. 2 (*water 3*) placed next to a water/vapor interface was spontaneously adsorbed at the interface. At the interface, the amount of β conformations decreased, and α-helical conformations formed in the C-terminal half of the peptide in two out of three simulations (compare Fig. 2 (*water/vapor*) and Fig. 1), suggesting early conformational transitions after adsorption. In all simulations of B18 at water/vapor interfaces, α-helical segments were approximately parallel to the interface.

4 Conclusion

Our results on the conformation of monomeric B18 in water and at a water/vapor interface and the interfacial orientation of the peptide are in agreement with available spectroscopic data and, thus, indicate these data to reflect properties of monomers. In addition, our simulations give insights into conformational distributions and transition pathways on a microscopic level. Revealing a pronounced conformational polymorphism of B18 in different environments, our work highlights the need for an explicit description of the solvent environment in order to understand fibrillogenic species from MD simulations.

References

1. A. S. Ulrich, M. Otter, C. G. Glabe, and D. Hoekstra, J. Biol. Chem. **273**, 16748 (1998).
2. R. W. Glaser, M. Grune, C. Wandelt, and A. S. Ulrich, Biochemistry **38**, 2560 (1999).
3. E. Maltseva, *Model membrane interactions with ions and peptides at the air/water interface*, Ph.D. thesis, Universität Potsdam (2005).
4. V. Knecht, H. Möhwald, and R. Lipowsky, J. Phys. Chem. B **111**, 4161 (2007).

Modeling the Free Energy of Polypeptides in Different Environments

Giovanni La Penna[1], **Sara Furlan**[1,2], **and Angelo Perico**[3]

[1] Institute for Chemistry of Organo–Metallic Compounds, National Research Council,
via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
E-mail: glapenna@iccom.cnr.it

[2] Department of Chemistry IFM, University of Torino,
via P. Giuria 7, 10125 Torino, Italy
E-mail: sfurlan@iccom.cnr.it

[3] Institute for Macromolecular Studies, National Research Council,
via De Marini 6, 16149 Genova, Italy
E-mail: perico@ge.ismac.cnr.it

The free energy for increasing the content of helical and elongated backbone segments in polypeptides is estimated performing computer simulations of "reasonable" random walks for all–atom models of isolated single chains. A summary of the applications on several homo polypeptides (X_{40} with X=G,A,V,T,K,E) and on the Aβ(1-40) peptide involved in Alzheimer disease is reported.

1 Introduction

The aggregation propensity, a consequence of the protein sequence and of its environment, can be the indirect effect of the absence of a folding propensity[1,2]. This effect has been called "inverse side chain effect"[3] or "natively disordered model"[2]: some sequences, when merged into a given environment, do not find a pathway to fold into protected structures and are therefore more suited to form a-specific backbone interactions that can eventually drive protein oligomerization.

We propose a model for measuring the competition between the formation of protected structures and structures suited to aggregation[4]. The propensity for a single chain a to adopt a given value, x, for a global configurational variable, $X(\mathbf{r})$ ($\mathbf{r}$ being the configuration of all atoms in the molecule), is measured by the free energy $f_a(x)$. In the model, collective properties that monitor the secondary structure of peptides were used as x. Because of the relevance of helical and elongated segments in the structure of peptides involved in formation of fibrils, we focused on the construction of extended helical segments or, alternatively, of extended elongated segments in different sequences of 40 aminoacids.

2 Method

Random walks were performed by using the Monte Carlo method in the dihedral space of single chains, with moves associated to randomly chosen temperatures (T_{max} =10000 K), the PARM99 force–field, a short cut–off for nonbonding interactions (0.5 nm) and biasing potential linearly dependent on the collective chosen variable X. The helical content was

measured via the maximal number of consecutive residues with $260^\circ < \phi < 320^\circ$ and $297^\circ < \psi < 353^\circ$ ($X = L_\alpha$). The β–strand content was measured as the maximal number of residues with $150^\circ < \phi, \psi < 210^\circ$ ($X = L_\beta$, hereafter).

The free energy $f_a(x) = u_a(x) - Ts_a(x)$ of chain a is given by the following equations. Assuming that the density of configurations is the average over all the biased metastatistics, the two contributions, s_a and u_a, to the free energy can be rewritten as:

$$\begin{aligned} s_a(x) &= R \log \left\{ \widetilde{n}'_a(x) \right\} + R \log \left\{ \langle \exp(V) \rangle'_a(x) \right\} \\ u_a(x) &= u_{0,a} + d_a RT/2 + \frac{\int_v U_a(\mathbf{r})\ \widetilde{P}'_a(\mathbf{r})\ \delta[X(\mathbf{r}) - x]\ d\mathbf{r}}{\widetilde{n}'_a(x)} \end{aligned} \tag{1}$$

where: $\widetilde{n}'$ and $\widetilde{P}'$ are, respectively, the density of x and the configurational probability evaluated collecting the metastatistics of the bundle of biased random walks; V is the biasing potential (linear functions of x); v is the space spanned by configurations $\mathbf{r}$; $X(\mathbf{r})$ is L_α or L_β (antagonist); f_a is the free energy in state x of chain a; u_a is its total energy in state x; $K_a = d_a RT/2$ is the kinetic part (with d_a torsional d.o.f.); $u_{0,a}$ is the reference state for energy ($L_{\beta,max}$).

The $U_a(\mathbf{r})$ term is given by the force–field, this time including long–range corrections, and adding mean–field corrections for the environment. To take into account water solvation, finite difference solutions of the Poisson-Boltzmann equation for the polypeptide in water, together with solvent accessible surface area contributions, were included in U_a.

3 Results

Homo polypeptides with sequences X_{40} (X=G,A,V,T,K,E) were studied. Their helical and β–strand propensity was compared with the Aβ(1-40) peptide involved in the Alzheimer disease and several peptides with the Aβ sequence randomly scrambled.

As a summary of this study, here we report in Fig. 1 the comparison between the free energy of the chosen variables for G_{40}, A_{40}, V_{40} and Aβ(1-40), in water solution (panel a) and in the vacuum (panel b), this latter modeling the membrane environment. The f profiles are shifted for graphical purposes.

Gly is unstructured in water solution, displaying only a moderate resistance towards extending its β–strand content. In a membrane–like environment, this resistance is almost completely lost and even small intermolecular interactions can easily stabilize extended β–sheets. Ala displays an opposite behavior: both in water solution and in the membrane–like environment the chain displays a high propensity of being structured in helical segments, while the resistance to extend β–strand content is large. Val displays an intermediate behavior: in water, the propensity for extending helical segments is similar to that of Gly, but the resistance to extend β–strands is similar to that of Ala; in the vacuum, the propensity for extending helical segments is similar to Ala. The Aβ(1-40) peptide displays a behavior that is similar to Val homo polypeptide. The significant propensity of extending helical segments in the membrane–like environment is expected, being Aβ(1-40) part of the membrane protein APP. But this propensity is almost entirely lost in water solution, where the $f(L_\alpha)$ profile is almost flat. On the other hand, no significant decrease in the resistance

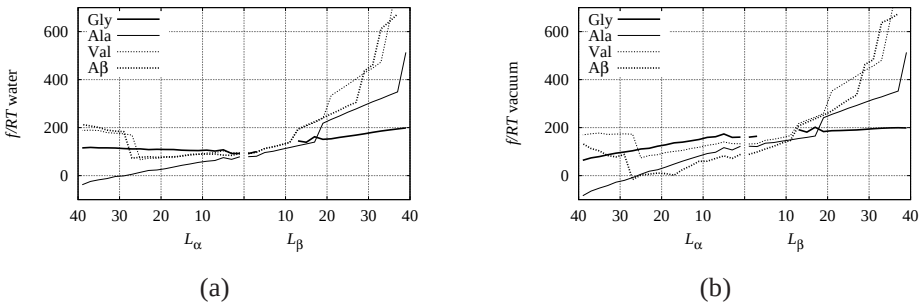


Figure 1. Comparison between free energy $f(L_\alpha)$ (left x-axis) and $f(L_\beta)$ (right x-axis) in Eq. 1 between several peptides, in water solution (a) and in the vacuum (b).

of extending β–strand is displayed when the peptide is extracted from the membrane–like environment into the water solution.

The similarity between Aβ(1-40) and Val_{40}, together with comparisons with other helical or β–strand propense peptides (data not shown here), implies that no significant propensity for extending β–strand length is encoded in the Aβ(1-40) sequence. Rather, a significant loss of propensity for extending helical content is displayed when the peptide is moved from the membrane to the water solution: aggregation is more the result of a low propensity for intramolecular folding than for a specific propensity for extending β–strands.

Acknowledgments

This work has been done with the financial support of the FIRB 2003 project RBNE03PX83 of MIUR (Italy).

References

1. C. M. Dobson, *Protein misfolding, evolution and disease*, TiBS, **24**, 329–332, 1999.
2. R. Nelson and D. Eisenberg, *Structural models of amyloid–like fibrils*, Adv. Prot. Chem., **73**, 235–282, 2006.
3. M. Fändrich and C. M. Dobson, *The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation*, EMBO J., **21**, 5682–5690, 2002.
4. S. Furlan, G. La Penna, and A. Perico, *Modeling the free energy of polypeptides in different environments*, Macromolecules, **41**, 2938–2948, 2008.

Protein-Ligand Docking with Two-Scale Receptor Dynamics and QM/MM Potential

Jarosław A. Kalinowski and Bogdan Lesyng

CoE BioExploratorium and Department of Biophysics, University of Warsaw,
Zwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: jak@biogeo.uw.edu.pl, lesyng@icm.edu.pl

A new docking model PhDock is presented. The docking process is correlated with global deformations (normal modes) of a receptor and with reorientations of its side chains. Deformations are derived from an elastic network model. Microscale motions of individual residues are generated using a statistical potential derived from a library of conformers. A basin-hopping Monte Carlo dynamics of a ligand is carried out using the SCC-DFTB energy plus protein - ligand and Lennard-Jones interactions. The electrostatic contribution is computed using SCC-DFTB atomic charges (Mulliken and CM3) in the presence of the receptor field, which accounts for the ligand polarization effects.

1 Introduction

Docking methods of small molecules to protein targets have been extensively developed in the past few decades. Due to constant increase of available computational power, some procedures that were previously out of reach become presently possible. We present a flexible-receptor and flexible-target model that is based on extensive two-scale probing of receptor configurations, and the ligand dynamics which is based on a QM/MM potential. The slow-motion, global receptor deformations, are accounted for by considering first few target normal modes[1] derived from an elastic network model[2]. In turn, faster reorientations of amino acid side chains allow achieving more energetically favorable interactions by the ligand. This reorientations are modelled by switching discreet side chain conformations, taken from a library[3], according to a Monte Carlo dynamics using library provided propensities as a stochastic potential. Mezoscopic and microscopic motions are decoupled, except for the checking of sterical conflicts. In the presence of the receptor with its own dynamics, a basin-hopping Monte Carlo ligand dynamics is carried out with the potential that combines ligand deformation energy computed using a fast SCC-DFTB QM method, together with protein-ligand coupling given by Coulomb and Lennard-Jones interactions with the receptor atoms. The electrostatic contribution is computed using ligand CM3/SCC-DFTB atomic charges[4] in the presence of the receptor, which accounts for electron polarization effects.

2 Potential Energy Function

The potential used for the ligand minimization and dynamics is defined as follows:

$$E = E^{LJ} + E^{ES}[q^{\mathrm{CM3+P}}; \epsilon] + E^{QM} \tag{1}$$

where E^{LJ} is the Lennard-Jones term computed using Amber94 parameters for the receptor, and a fixed set of LJ parameters for the ligand atoms. The remaining terms represent

the Coulomb interaction energy with the CM3 charges and polarization corrections to the ligand charges, as well as the ligand quantum mechanical energy, which is computed using the SCC-DFTB method, which has an algebraic structure of a Tight-Binding method with the energy decomposition into the "band structure" and "repulsion":

$$E = \sum_{i}^{\text{occ.}} \epsilon_i + \frac{1}{2} \sum_{\text{A}\neq\text{B}} U_{\text{AB}}(|R_{\text{A}} - R_{\text{B}}|). \tag{2}$$

The Hamiltonian, however, has an extra term $H^1_{\mu\nu}$ which includes interactions of intramolecular partial charges and interaction with external electrostatic field:

$$H^{\text{SCC-DFTB}}_{\mu\nu} := H^{\text{DFTB}}_{\mu\nu} + H^1_{\mu\nu} \qquad H^1_{\mu\nu} = S_{\mu\nu}(\Gamma_{\text{A}} + \Gamma^{\text{ext.}}_{\text{A}}), \;\; \nu \in \text{A} \tag{3}$$

where Γ_{A} and $\Gamma^{\text{ext.}}_{\text{A}}$ are corresponding potential shifts on atom A, and $S_{\mu\nu}$ is the overlap matrix. Atomic charges on the receptor are included in ligand SCC-DFTB calculations *via* $\Gamma^{\text{ext.}}_{\text{A}}$ shifts:

$$\Gamma^{\text{ext.}}_{\text{A}} = \sum_{\text{n}} \frac{1}{\epsilon} \frac{Q_{\text{n}}}{|R_{\text{A}} - R_{\text{n}}|} \tag{4}$$

where Q_{n} are the receptor charges, and ϵ scales the strength of the interaction potential.

2.1 CM3/SCC-DFTB Charges

The CM3 procedure computes a correction to the Mulliken charges after the SCC cycle, which reduces systematic errors of individual bond dipoles:

$$q_k = q^0_k + \sum_{k'\neq k} T_{kk'}(B_{kk'}) \tag{5}$$

where q_k is the CM3 charge on an atom k, q^0_k is the original Mulliken charge, $B_{kk'}$ is the Mayer bond order and $T_{kk'}$ is a function of the bond orders which determines the amount of the charge to be transferred from an atom k' to an atom k:

$$T_{kk'} = D_{Z_k Z_{k'}} B_{kk'} + C_{Z_k Z_{k'}} (B_{kk'})^2. \tag{6}$$

The C and D coefficients were determined by the parameterization procedure.

3 Sampling Scheme

3.1 Side Chain Conformer Libraries

We use libraries constructed and published by Shetty et al.[3]. These side chain conformer libraries were extracted from high-quality protein structures, they maintain crystallographic bond lengths and angles, in contrast to traditional rotamer libraries defined in terms of χ angles under the assumption of idealized covalent geometry. The libraries provide also backbone ϕ, ψ dependent propensities for each 40° dihedral angle bin, computed from the conformational populations.

The conventional Metropolis algorithm[5] is formulated as a rule which states that transition from a state A to a state B is accepted with the probability $\exp(-(E_B - E_A)/kT)$. This is, however, equal to the ratio P_A/P_B of the probabilities of states A and B. In this approach the probabilities are taken from the normalized propensities included in the SCL libraries.

Destination conformations are drawn with equal probabilities from the whole ensemble of conformations allowed by the ϕ, ψ angles. Frequency of side chain orientation changes for a given residue is determined by its distance to the ligand. Currently a dumping function $e^{-(\frac{r}{r_0})^2}$ is used with r_0 in range of 0.7...1.2 nm.

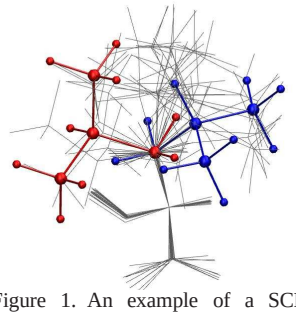


Figure 1. An example of a SCL step. Two randomly selected conformations are shown using the CPK model. Every second conformation from the SCL0.5 library is drawn using thin lines.

3.2 Normal Modes

Normal modes are computed using MMTK library[6] written by Konrad Hinsen, in the C_α representation of the protein. The applied potential is defined as follows:

$$V(\vec{r}_1 \dots \vec{r}_N) = \sum_{i<j} f(r_{ij}^0)(r_{ij} - r_{ij}^0)^2$$

where $r_{ij} = |\vec{r}_i - \vec{r}_j|$, and the pair force constant $f(r_{ij}^0)$ is given by the expression:[2]

$$k(r_{ij}^0) = \begin{cases} 8.6 \times 10^5 r_{ij}^0 - 2.39 \times 10^5 & r_{ij}^0 < 0.4\text{nm} \\ 128(r_{ij}^0)^{-6} & r_{ij}^0 \geq 0.4\text{nm} \end{cases}$$

using nm and kJ/mol units, and depends on residue distance in the reference conformation. Diagonalization of the second derivatives matrix using a fourier basis gives the modes $m_{\mu i}$ and the corresponding frequencies f_μ. The resulting atomic displacement vectors for selected modes (default: the lowest four nonzero modes) are copied from the C_α atoms to all remaining atoms in each residue.

Let the nm-state be defined as $d = (d_1 \dots d_m) \in \mathbb{R}^m$, where m is the number of selected modes and d_μ can be interpreted as a measure of deformation along the μ-th selected mode. The receptor atom i displacement is given by:

$$\vec{r}_i = \vec{r}_i^{\,0} + \sum_\mu d_\mu m_{\mu i}$$

The Monte Carlo dynamics using the Metropolis algorithm[5] is carried out with the quadratic potential $V(d) = \sum f_\mu d_\mu^2$. The temperature does not have any well defined physical meaning and is treated as a parameter of the model.

3.3 Algorithm

The algorithm consists of a sequence of steps that can be summarized as follows:

draw a step-type with probabilities: *nm:scl:rb:ligmin* (typically: 1:4:64:1 or 2:4:32:1)
if step-type is *nm* **then**
  $d^{\text{new}} = d + \delta$ where δ is a random vector in $\mathbb{R}^m$
  if step valid and accepted by `Metropolis`(E^{new}, E) **then**
    apply deformation resulting from δ to the receptor; find new basin minimum
  end if
else if step-type is *scl* **then**
  draw residue using probabilities given by the dumping function $\exp(-r^2/r_0^2)$
  draw new conformation from conformations allowed by residue ϕ, ψ angles
  if step valid and accepted with probability P_A/P_B **then**

change side chain orientation; find new basin minimum
end if
else if step-type is *rb* **then**
draw transformation being superposition of random, rigid rotation and translation
find new ligand configuration and its basin minimum $E_{\text{basin}}^{\text{new}}$
if step valid and accepted with probability $\texttt{Metropolis}(E_{\text{basin}}^{\text{new}}, E_{\text{basin}})$ **then**
accept new ligand position and new basin minimum
end if
else if step-type is *ligmin* **then**
perform fully flexible ligand minimization
end if

It should be noted the energy minimization, as a part of the basin-hopping scheme, is unrestricted allowing flexible deformations.

4 Results and Conclusion

The novel, promising multiscale QM/MM flexible docking method (PhDock) was formulated, implemented, and is being tested using a number of biomolecular model systems. These, in particular, include a short helix with several water molecules and three protein-ligand complexes with PDB codes: 1ctt, 1ivd and 2qwe. The applied stochastic sampling procedure located the correct binding modes for the first two complexes. In the last, difficult case with a deeply buried ligand, the sampling procedure located a few alternative binding poses on the protein surface. Longer sampling procedures along with refinement of selected control parameters of the model are being carried out.

Acknowledgments

These studies are supported by the N207 022 31/1172 grant and by the Centre of Excellence BioExploratorium.

References

1. M. Zacharias *Rapid Protein-Ligand Docking Using Soft Modes From Molecular Dynamics Simulations to Account for Protein Deformability: Binding of FK506 to FKBP* Proteins **54**, 759–767, 2004.
2. K. Hinsen, A.J. Petrescu, S. Dellerue, M.C. Bellissent-Funel, G. R. Kneller *Harmonicity in slow protein dynamics* Proteins **261**, 25-37, 2000.
3. R. P. Shetty, P. I.W.de Bakker, M.A. DePristo, T.L. Blundell *Advantages of fine-grained side chain conformer libraries* Protein Engineering **16**, 963–969, 2003.
4. J.A. Kalinowski, B. Lesyng, J.D. Thompson, C.J. Cramer, D.G. Truhlar *Class IV Charge Model for the Self-Consistent Charge Density-Functional Tight-Binding Method* J. Phys. Chem. A **108**, 2545–2549, 2004.
5. Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller *Equation of State Calculations by Fast Computing Machines* J. Chem. Phys. **21**, 1087–1092, 1953.
6. K. Hinsen *The Molecular Modeling Toolkit: A New Approach to Molecular Simulations* J. Comp. Chem. **21**, 79–85, 2000.

Inhibition of Fibril Formation of Beta-Amyloid Peptides

Nguyen Sy Lam[1], Maksim Kouza[2], Hoang Zung[1], and Mai Suan Li[2]

[1] Computational Physics Lab, Vietnam National University,
Ho Chi Minh city, 227 Nguyen Van Cu, Dist. 5, Vietnam

[2] Institute of Physics, Polish Academy of Sciences,
Al. Lotnikow 32/46, 02-668 Warsaw, Poland
E-mail: masli@ifpan.edu.pl

Alzheimer's, Parkinson's, Huntington's, type II diabetes, and Mad Cow disease, or Cystic fibrosis, these apparently unrelated diseases, the so-called protein structural diseases, are found to be a result of protein misfolding. Understanding the role of molecular inhibitors in formation of amyloid fibrils plays an important role in finding proper treatments to those structural diseases. In the case of Alzheimer's disease, experiments showed that the fibrillation of full-length $A\beta$ peptides is disrupted by the peptide fragment $A\beta_{16-20}$ (KLVFF). In this contribution, we studied the kinetics of oligomerization of the system of two $A\beta_{16-22}$ and one $A\beta_{16-20}$ peptides, using all-atom simulations with the GROMOS96 force field 43a1 in explicit water. In agreement with experiments, $A\beta_{16-20}$ peptide was found to slow down the aggregation process.

1 Introduction

In many cases protein aggregates take the form of amyloid fibrils, which appear as unbranched rod-like nanostructures with the diameter of an order of 10 nm and varying length[1]. A large body of evidence suggests that amyloid fibrils and associated oligomeric intermediates are related to a number of diseases, including Alzheimer's, Parkinson's, Huntington's, and prion diseases[1]. For example, in the case of the Alzheimer's disease the memory decline may result from the accumulation of the amyloid β-proteins (Aβ) present in two forms - 40 (Aβ_{1-40}) and 42 (Aβ_{1-42}) amino acids of which are produced through endoproteolysis of the β-amyloid precursor transmembrane protein. Since structural diseases affect a significant portion of senior population, it is vital to develop therapeutic approaches to combat the amyloid assembly. One of possible ways is to design molecular inhibitors, which interfere with this process. For example, the peptide fragment Aβ_{16-20} KLVFF[2] and the peptide LPFFD derived from Aβ_{17-21} fragment by V18P and A21D mutations[3] can disrupt fibrillation of full-length Aβ peptide. An insertion of prolines also inhibits amyloid formation. An another powerful strategy for inhibition is a peptide N-methylation. It was demonstrated that themembrane-permeable NN-methylated pentapeptide A$\beta 16-20$m is an effective fibrillogenesis inhibitor, capable of both preventing fibril growth and disassembling existing fibrils[4]

In this contribution, we study the influence of a pentapeptide Aβ_{16-20} on the kinetics of oligomerization of longer Aβ_{16-22} peptides, using all-atom simulations with the GROMOS96 force field 43a1 in explicit water[5]. Since the fibril formation time t_{fibril} of a system of two Aβ_{16-22} and one Aβ_{16-20} peptide is much longer than that for three Aβ_{16-22} peptides, one can expect that, in agreement with experiments[2], Aβ_{16-20} fragment slows down the fibril growth process.

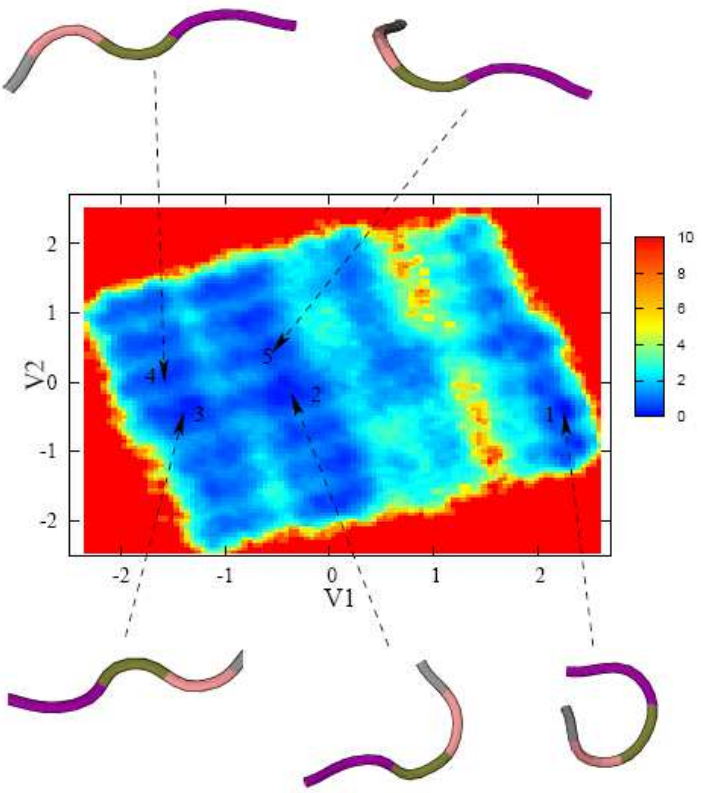


Figure 1. The free energy landscape as a function of V_1 and V_2. Typical conformations of some local minima are shown.

2 Method and Results

The structures of monomeric Aβ_{16-20} and Aβ_{16-22} were extracted from the structure of Aβ_{10-35} peptide available in the Protein Data Bank (ID: 1hz3). In order to study conformation changes of a monomer Aβ_{16-20}, one trajectory of 150 ns was generated. The initial configuration of the system of two Aβ_{16-22} and one Aβ_{16-20} was created by randomly placing these peptides in a periodic box of volume 78 nm^3 which corresponds to the peptide concentration of 64 mM. For this system, four runs of 300, 343, 453, and 484 ns were carried out. All simulations were performed at $T = 300$ K.

We used the dihedral principal component analysis[6] to compute the free energy landscapes (FEL), using the first two eigenvectors V_1 and V_2. In order to monitor the fibril formation process, we use the "liquid crystal" order parameter P_2. If $P \geq 0.9$ then the system is considered to be in the fibril-like state[6].

Fig. 1 shows the free energy of a monomer Aβ_{16-20} as a function of V_1 and V_2. The existence of many shallow local minima separated by low barriers (of only a few k_BT)

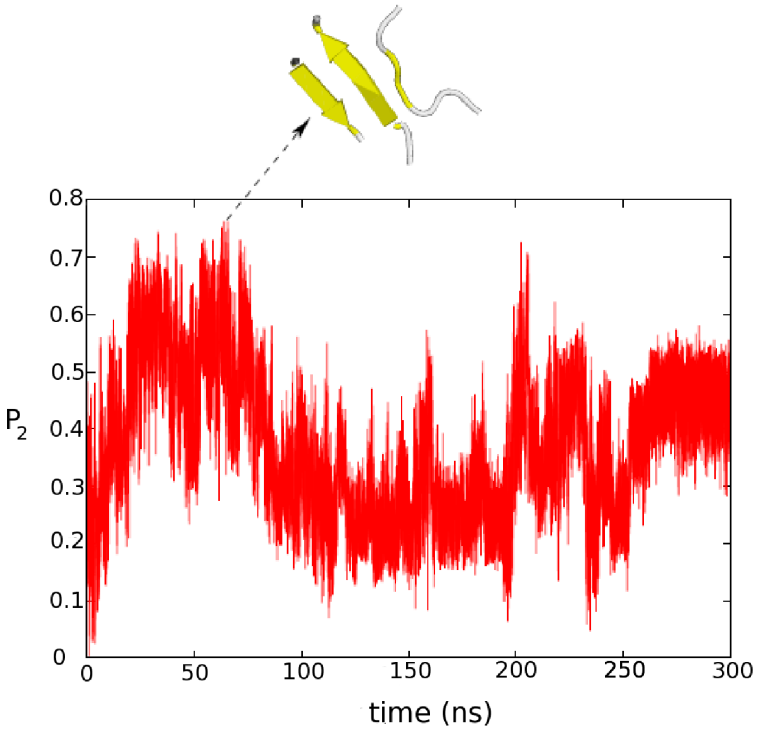


Figure 2. Time dependence of P_2 for the (2Aβ_{16-22}+Aβ_{16-20}) system. Shown is the structure at the largest $P_2 \approx 0.75$. Interestingly, one Aβ_{16-22} and one Aβ_{16-20} adopt anti-parallel arrangement.

suggests that the system is not stable under thermal fluctuations. As evident from typical snapshots (Fig. 1), the monomer adopts mainly coil conformations. Comparing with the free energy landscape of Aβ_{16-22}[6], one can see that the pentapeptide Aβ_{16-20} is less stable than Aβ_{16-22}. This may be a reason why Aβ_{16-20} can serve as an inhibitor for oligomerization of Aβ_{16-22} peptides, because in some situations proteins which have a more ordered structure in the monomeric state, are expected to be more prone to aggregation.

The time dependence of the order parameter P_2 for the (2Aβ_{16-22}+Aβ_{16-20}) system is shown in Fig. 2. Since the largest value $P_2 \approx 0.76$, found at $t \approx 64$ ns, is lower than 0.9, a fibril-like state does not occur in this run. This state was also not observed for three other trajectories (results not shown). Therefore, for the (2Aβ_{16-22}+Aβ_{16-20}) system, $t_{fibril} > 400$ ns, which is larger than $t_{fibril} \approx 200$ ns for a system of three Aβ_{16-22} peptides[6]. This result suggests that, in agreement with the experiment[2], Aβ_{16-20} can interfere with the oligomerization process. There are two possible reasons for this:
(a) Aβ_{16-20} does not contain the negatively charged glutamic acid (E) as Aβ_{16-22} does.
(b) The replacement of Aβ_{16-22} by KLVFF reduces hydrophobicity of the whole system.

In conclusion, we have shown that a short peptide KLVFF can inhibit the oligomerization of a system of Aβ_{16-22} peptides. It is expected to prevent the fibril growth of full-length $A\beta$ peptides, due to charge imbalance.

Acknowledgments

This work was supported by the Ministry of Science and Informatics in Poland (grant No 202-204-234).

References

1. F. Chiti, and C. M. Dobson, *Protein misfolding, functional amyloid, and human disease*, Annual Rev. Biochemistry **75**, 333–366, 2006.
2. L. O. Tjernberg *et al.*, *Arrest of β-amyloid fibril formation by a pentapeptide ligand*, J. Biol. Chem. **271**, 8545-8548, 1996.
3. C. M. Soto, S. Kindy, M. Baumann, and B. Frangione, *Inhibition of Alzheimer's amyloidosis by peptides that prevent β-sheet conformation*,Biochem. Biophys. Res. Commun. **226**, 672-680, 1996.
4. D. J. Gordon, R. Tape, and S. C. Meredith, *Design and characterization of a membrane permeable N-methyl amico acid-containing peptide that inhibit Aβ_{1-40} fibrilogenesis*, J. Pep. Res. **60**, 37-55, 2003.
5. H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *A message-passing parallel molecular dynamics implementation*, Comp. Phys. Comm. **91**, 43–56, 1995.
6. P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub, and D. Thirumalai, *Monomer adds to preformed structured oligomers of Aβ-peptides by a two-stage dock-lock mechanism*, Proc. Natl. Acad. Sci. USA **104**, 111–116, 2007.

A Topological Model of a "Jumping Gene" Machine

Junalyn Navarra-Madsen

Department of Mathematics and Computer Science,
Texas Woman's University, Denton TX 76210, USA
E-mail: jnavarramadsen@mail.twu.edu

Mu is similar to a human immunodeficiency virus (HIV) capable of transposing or integrating itself into the host genome. This paper focuses on a topological model describing the changes in 2-dimensional conformation of the DNA upon Mu binding and catalysis.

1 Introduction

Most of the time, due to the difficulty of making *in vivo* experiments, molecular biologists start with an *in vitro* system of a certain biological process they want to understand. One such important process is DNA transposition. Mu virus was the first transposition system studied in vitro[2]. This in vitro transposition system aided in understanding the structure and function of transposition proteins in nucleoprotein complexes called transpososomes[2,1]. The protein-DNA core of the Mu transpososome is composed of a tetramer of the transposase, Mu A, bound to the two special DNA ends called transposons, sometimes referred to as mobile elements. The reasons for constructing a topological model of Mu-DNA complex are two-fold: first, having an in vitro system is not good enough since the process of isolation and storage of a cellular system may change part or the whole system itself; second, there is no known high-resolution structure of the Mu transpososome to date. A mathematical model (topological or geometric) can be helpful in removing redundancies thereby reducing the number of biologically plausible cases and saving valuable time.

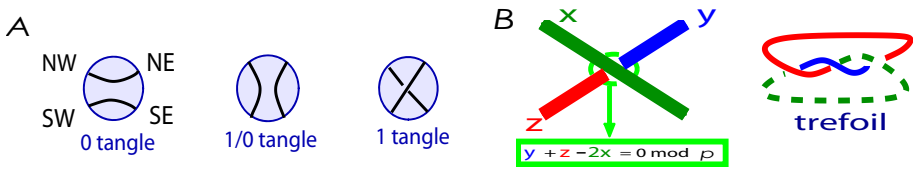


Figure 1. A) Examples of 2-string tangles; B) Definition of colorability. A coloring of diagram of a knot or link or tangle is a function $C : \{arcs\ of\ a\ diagram\} \mapsto \mathbb{Z}_p$ such that at each crossing the relation $y + z - 2x = 0$ mod p holds, where x is the color on the over arc and y and z are the two colors of two under arcs. A three-crossing knot or a trefoil is 3-colorable.

2 Biological Background

In order to topologically study the mechanism of a certain enzyme, molecular biologists start with a circular DNA. Conformational changes on linear DNA made by the enzyme

during catalysis can easily slip and be lost. If one starts with an unknotted double-stranded DNA and obtains a knotted circular DNA, then the enzyme must have acted on the DNA. One possible sequence of enzyme action is: cut, exchange and reseal two DNA segments. Sumners, Ernst and collaborators showed the first mathematical model of DNA recombination using tangles[3]. A n-string tangle is a set of n strings properly embedded in a 3-dimensional ball, Fig. 1A. For a thorough review of tangles and tangle calculus, see a couple of papers by Ernst[5,6]. The 3-dimensional ball represents the protein and the supercoiled string represents double-stranded DNA. Pathania *et al* used circular DNA with properly placed recombination (lox P) and transposition (attL and attR) sites so that they could observe specific changes in these special DNA sites upon protein binding and catalysis[1]. See Fig. 2A. When translated to tangles' parlance, there are two tangles involved, the tangle modeling a better-understood process of Cre recombination[7] and the tangle describing Mu transposition. In the lower right hand corner of Fig. 2B, Cre recombinase changes a 0-tangle to a $\frac{1}{0}$-tangle. This Cre recombination tangle was used as a tool to understand what DNA conformation was trapped in the Mu transpososome upon Mu binding, and a 2-dimensional solution (a tangle with five crossings) obtained by Pathania *et al* after performing all the experiments with different maps where they assumed that the DNA trapped during Mu transposition is *branched supercoiled*, and an example of a branched supercoiled tangle. Most mathematicians are concerned with the question of uniqueness of solution. In the following section, one topological confirmation of uniqueness of this five-crossing solution presented by the in vitro experiments of Pathania *et al* is shown. One strength of this topological modeling is that there is no assumption made about the transposition tangle.

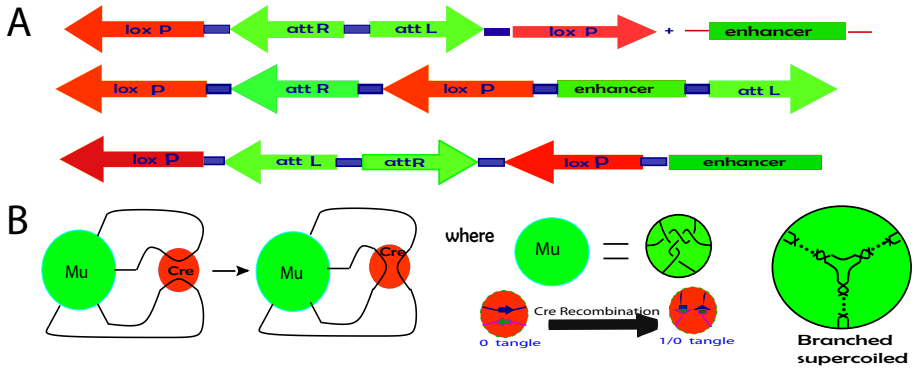


Figure 2. A) Different kinds of maps of Cre recombination and Mu transposition sites; B) 2-dimensional model of the 3-dimensional Mu-DNA transpososome. Pathania *et al* obtained this five-crossing model of Mu-DNA complex[1].

3 Topological Invariant: Colorability

One can use a knot invariant called colorability (Fig. 1B) to mathematically describe the DNA conformational changes formed after the protein binds the DNA and has acted on

it. Darcy *et al* proposed a computational method using colorability to encode DNA conformational changes[8]. Every knot diagram **K** with k crossings has exactly k arcs. By the definition of colorability, there is a $k \times k$ matrix corresponding to a knot diagram **K** with k crossings and k arcs. Therefore, **K** can be colored mod p if and only if the corresponding set of equations has a nontrivial mod p solution[9]. Similarly, tangle diagrams can be colored, Fig. 3A.

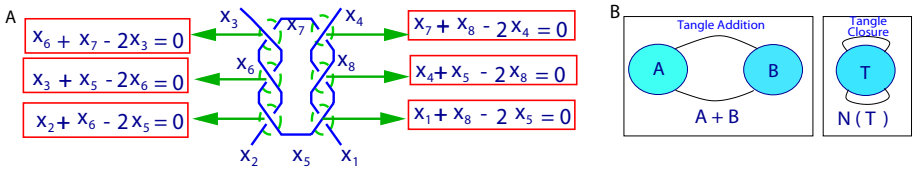


Figure 3. A) An example of coloring a 2-string tangle. There are six crossings (equations) and eight arcs (variables). In[10], it was shown that two tangle invariants can be used to distinguish one n-string tangle from another; B) Tangle addition and tangle closure.

Tangles can be embedded in a knot. Using tangle calculus, a knot can be written as a numerator closure of the sum of two tangles, Fig. 3B. This can then be written as a tangle equation. Two examples of tangle equation is in Fig. 4. One starts with a circular unknot (no crossings when simplified topologically) and ends up with a three crossing knot. These two tangle equations can then be translated into colorability as systems of equations solvable modulo p, $p \in \mathbb{Z}$, given crossings of tangles involved.

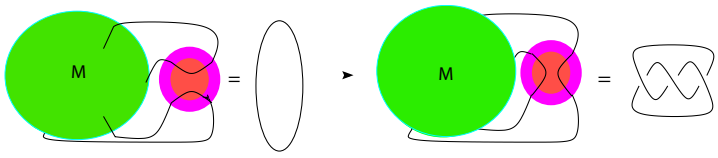


Figure 4. Two examples of tangle equations, one starts with an unknot and ends with a three-crossing knot.

To make the long story short, with the use of two simple tangle invariants found in Ref. 10 and some minor computation, there is one and only one solution to the system of tangle equations, thereby confirming the in vitro solution of Pathania *et al*.

4 Concluding Remarks

One question arises from this study. If given the right length of DNA, will the Mu transpososome take the configurations of crossings greater than five given in the summary table, Table 1.

No. of Crossings	Correct Coloring Matrices	Non-equivalent and Colorable
≤ 4	0	0
5	1	1
6	22	0
7	354	3
8	5019	6

Table 1. Summary table of tangles given the number of crossings in the Mu 2-dimensional configurations.

Acknowledgments

Thank you to the Chancellor's Research Fellows Program of Texas Woman's University for some research support.

References

1. S. Pathania, M. Jayaram, and R. Harshey, *Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils*, Cell **109**, 425–436, 2002.
2. G. Chaconas and R. Harshey, *Transposition of phage Mu DNA, Mobile DNA: Volume II*, (ASM Press, 1752 N Street, NW Washington, DC, 2002).
3. D. W. Sumners, C. Ernst, S. Spengler and N. Cozzarelli, *Analysis of the mechanism of DNA recombination using tangles*, Quarterly Reviews of Biophysics **28**, 253–313, 1995.
4. C. Ernst and D. W. Sumners, *A calculus for rational tangles: applications to DNA recombination*, Math Proc Cambridge Philos Soc **108**, 489–515, 1990.
5. C. Ernst, *Tangle equations*, J Knot Theory Ramifications **5:2**, 145–159, 1996.
6. C. Ernst, *Tangle equations II*, J Knot Theory Ramifications **6**, 1–11, 1997.
7. A. A. Vetcher, A. Y. Lushnikov, J. Navarra-Madsen, R. G. Scharein, Y. L. Lyubchenko, I. K. Darcy, and S. D. Levene, *DNA Topology and Geometry in Flp and Cre Recombination*, J Mol Biol **357**, 1089–1104, 2006.
8. I. K. Darcy, J. Chang, N. Druivenga, C. McKinney, R. Medikonduri, S. Mills, J. Navarra-Madsen, A. Ponnusamy, J. Sweet, and T. Thompson, *Coloring the Mu transpososome*, BMC Bioinformatics **7**, 435–454, 2006, `http://www.biomedcentral.com/1471-2105/7/435`.
9. R. H. Fox, *A Quick Trip Through Knot Theory*, in Topology of 3-manifolds and related topics (edited by M. K. Fort), 120–167, 1961.
10. J. Navarra-Madsen and I. K. Darcy, *Coloring n-string Tangles*, submitted to Journal of Knot Theory and its Ramifications, 2008.

Computer Modeling of Small Ligands Diffusion in Drosophila Melanogaster Hemoglobin

Lukasz Dams, Slawomir Orlowski, and Wieslaw Nowak

Institute of Physics, N. Copernicus University,
ul. Grudziadzka 5, 87-100 Torun, Poland
E-mail: wiesiek@fizyka.umk.pl

The monomeric, intracellular hemoglobin from the fruit fly Drosophila Melanogaster (DmHb) has been discovered in 2005. It came out that the oxygen supply system in insects is more complex then previously thought. Details on diffusion of gases, ligands discrimination and hexa- to pentacoordination changes remain unclear. Here we present the results of molecular modeling and molecular dynamics (MD) simulations of gaseous ligands (O_2, CO, NO) transport inside the DmHb matrix. In addition to a classical MD trajectory an approximate Locally Enhanced Sampling method (LES) and Implicit Ligand Sampling (ILS) have been employed. The structural and thermodynamics features of hexacoordinated and pentacoordinated DmHb were examined and compared to our previous results obtained for human neuroglobin and cytoglobin which display similar heme coordination. Several connected cavities and diffusion pathways, based on 3D ILS free-energy maps, have been indicated. Residues that are critical for kinetics of small gaseous ligands diffusion in primitive hemoglobin were discovered. These data may help do understand the impact of evolutionary pressure on proteins architecture.

1 Introduction

Recently, *Drosophila melanogaster* Hemoglobin (DmHb), a new member of invertebrate heme globin family, has been discovered[1]. It belongs to the class of hexacoordinated globins The sixth coordination position of heme iron ion is occupied by an external ligand (i.e. O_2, CO) or the distal HisE7. *D. melanogaster* has the open circulatory system with fluid called hemolyph. In hemolyph there are no hemoglobin-like proteins. Tracheal system supports passive diffusion of oxygen to the tissues and for many years researchers thought that oxygen carriers are unnecessary in this taxon. DmHb expresses in fat body and pharynx muscle, so it is an intracellular protein. Its role is unclear and is a matter of the debate. Our goal is to find cavities network and diffusions pathways for small gaseous ligands. These data may help to understand the physiological role of DmHb and bring new information on the architecture and function of this ancient protein.

2 Methods

The structure of DmHb (2G3H) was obtained from Protein Data Bank. All MD simulations were carried out with the NAMD 2.6 code and CHARMM27 force field[2]. In the DmHb structure heme group coordinates cyanide ligand. To performe MD simulations with dioxygen, the cyanide has been removed and dioxygen ligand has been docked. This initial structure was placed in an equilibrated TIP3 water box (63x54x45 $Å^3$). The relaxation, heating to 300 K, and equilibration of the system (about 17 000 atoms) were completed during a 500 ps run. A standard MD trajectory 15 ns long was produced. Additionally, the

Locally Enhanced Sampling (LES) method was used to facilitate O_2 diffusion[3]. The three 3 ns long LES trajectories with 5, 10 and 15 copies of O_2 for DmHb structures were generated (named LES5, LES10 and LES15, respectively). These simulations were run with the periodic boundary conditions. The Ewald mesh summation for long range electrostatic interactions was used. The integration timestep was 1 fs for bonded and non-bonded interactions and 2 fs for electrostatics. The cutoffs for Van der Waals and electrostatic interactions were 12 Å. Langevin dynamics with dumping factor 5 $ps^{-}1$ was used. The Potential of Mean Force by Implicit Ligand Sampling (PMF/ILS) calculation was carried out with the VMD code[4]. This method allows for indication of all low free-energy profile within a protein matrix. The analysis was performed using the VMD code[5].

3 Results and Discussion

The inspection of root-mean-square (rms) distances from the DmHb model shows that all trajectories are reasonable stable (data not shown, average Calpha rmsd < 1.5 Å). In a static crystal structure DmHb no entry channels from the solvent to the heme active site can be found[1]. Moreover only three cavities are observed. However, when fluctuations of amino acids side chains at 300K were added, the transient routes leading from the solution to the binding site are presented. Using the PMF/ILS method such paths were determined for 3 small gaseous ligands: NO, CO and O_2. The complex network of cavities and channels in DmHb are determined by tracking dioxygen collisions with residues (LES trajectories) and by inspection of the PMF isosurface for selected ligands. Detailed data about amino acids involved in formation main cavities may be found in Fig. 1. Using the VMD soft-

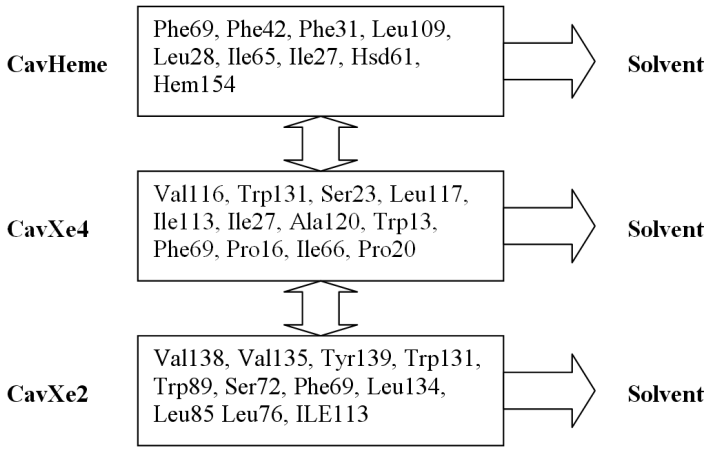


Figure 1. Diagram of the cavities in DmHb.

ware, traces of every O_2 ligand were registered in all calculated LES trajectories. Out of 5+10+15 ligands observed only 12 copies left the protein on a 3 ns timescale. Cavities DP, Xe4 and Xe2 exhibit exits path to the solvent. PMF isosurface is shown in the Fig. 2.

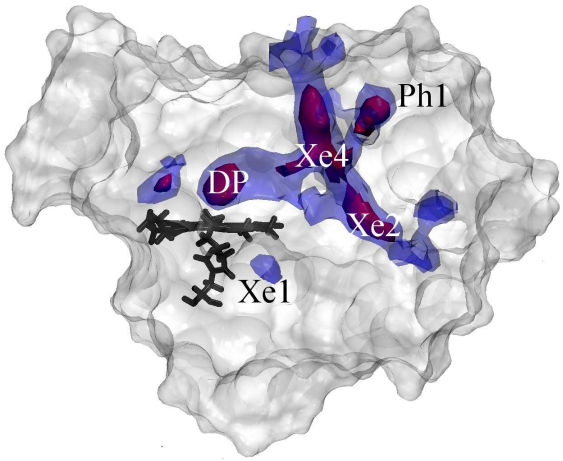


Figure 2. Free energy map for O_2 migration inside DmHb. Two free energy isosurfaces represent PMF/ ILS values of -1.5 kcal/mol (opaque black) and 1.6 kcal/mol (transparent dark gray). The DmHb surface is displayed in light gray and the heme group is represented in dark gray .

Five cavities have been found. Three of them are involved in building a big, hydrophobic channel (DP, Xe4 and Xe2). Cavities Xe1 and Ph1 are not connected with the channel thus their influence on the ligand diffusion may by neglected. The Ph1 pocket has connection only with the surface of the protein. These data are different from the standard myoglobin picture and are close to our previous study for the neuroglobin, cytoglobin and mini-hemoglobin[6,7]. Maybe hexacoordination provide a completely different mechanism for controlling ligand affinity and tune ligand diffusion rate. Within the hydrophobic channel at least 4 different exits are located. From the distal pocket a gaseous ligand can move to Xe4 cavity or directly to the solvent passing HisE7 residues. From Xe4 cavities ligand travels either to Xe2 or escape to the solvent using rather big exit placed between AB turn and E helix. Another exit path from Xe2 pocket is located between G and H helices. Ligand trapped in Xe2 pocket can move to solution between EF turn and H helix. This is another large tunnel leading to exterior.

4 Conclusions

Our dynamical data provide new information in ccontrast to the static analysis of one X-ray structure. The thermodynamic PMF/ILS calculations confirmed the existence of multiple diffusion paths in DmHb. These paths are easily accessible to small gaseous ligands. Despite of well conserved tertiary structure, globins have very different pathways for ligand diffusion. Composition of the amino acids may be more important in this case than the tertiary structure. Present data should contribute to evolutionary studies of oxygen transport mechanism in biological systems.

References

1. D. Plimpton, *et al*., *Bishistidyl heme hexacoordination, a key structural property in Drosophila melanogaster hemoglobin*, J. Biol. Chem. **280**, 27222–27229, 2005.
2. J.C. Phillips, *et al*., *Scalable molecular dynamics with NAMD* J. Comp. Chem. **26**, 1781, 2005.
3. R. Elber, M. Karplus, *Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin*, J. Am. Chem. Soc. **112**, 9161, 1990.
4. J. Cohen, *et al*., *Imaging the migration pathways for O2, CO, NO, and Xe inside myoglobin*, Biophys. J. **91**, 1844–1857, 2006.
5. W. Humphrey, A. Dalke, K. Schulten, *VMD - Visual Molecular Dynamics*, J. Molec. Graphics. **14**, 33, 1996.
6. R. F. Tilton, I. D. Kuntz,G. A. Petsko, *Cavities in proteins: structure of a metmyoglobinxenon complex solved to 1.9 Å*, Biochemistry **23**, 2849–2857, 1984.
7. S. Orlowski, W. Nowak, *Oxygen diffusion in minihemoglobin from Cerebratulus lacteus: a locally enhanced sampling study*, Theor. Chem. Acc. **117**, 253-258, 2007.

High Throughput In-Silico Screening Against Flexible Protein Receptors

Horacio Sánchez, Bernhard Fischer, Holger Merlitz, and Wolfgang Wenzel

Institut für Nanotechnologie, Forschungszentrum Karlsruhe,
76344 Eggenstein-Leopoldshafen, Germany
E-mail: horacio.sanchez@int.fzk.de

Results for the screening of known ligand databases versus rigid and flexible receptors are presented using an all-atom model and a very efficient optimization method. The results are compared with other programs and a better performance is shown.

1 Introduction

Virtual screening of chemical databases to targets of known threedimensional structure is developing into an increasingly reliable method for finding new lead candidates in drug development. Based on the stochastic tunneling method (STUN[1]) we have developed FlexScreen[2,3], a novel strategy for high-throughput in-silico screening of large ligand databases. Each ligand of the database is docked against the receptor using an all-atom representation of both ligand and receptor. In the docking process both ligand and receptor can change their conformation. The ligands with the best evaluated affinity are selected as lead candidates for drug development. Using the thymidine kinase inhibitors as a prototypical example we documented the shortcomings of rigid receptor screens in a realistic system. We demonstrate a gain in both overall binding energy and overall rank of the known substrates when two screens with a rigid and flexible (up to 15 sidechain dihedral angles) receptor are compared. We note that the STUN suffers only a comparatively small loss of efficiency when an increasing number of receptor degrees of freedom is considered. FlexScreen thus offers a viable compromise between docking flexibility and computational efficiency to perform fully automated database screens on hundreds of thousands of ligands.

2 Methodology

Docking Method: Stochastic optimization with STUN: Non-linear transformation of the potential energy surface using

$$E_{\mathrm{STUN}}(x) = \ln\left(x + \sqrt{x^2+1}\right) \ , \tag{1}$$

with $x = \gamma\,(E - E_0)$, $\gamma = 0.05$ Mol/kJ and E_0 is the lowest energy encountered during the simulation.

Scoring Function:

$$S = \sum_{Protein} \sum_{Lig.,fl.SC.} \left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}}\right) + \sum_{h-bonds} \cos\,\Theta_{ij} \left(\frac{\tilde{R}_{ij}}{r_{ij}^{12}} - \frac{\tilde{A}_{ij}}{r_{ij}^{10}}\right) \tag{2}$$

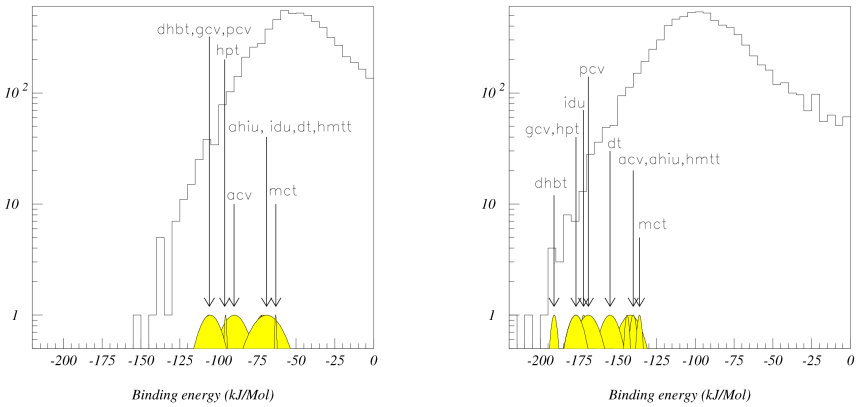


Figure 1. Binding energies of the docked substrates. Left side, rigid screening with total score 4206. Right side, flexible screening with total score 8083.

Partial charges q_i are usually evaluated with InsightII and ESFF forcefield, Lennard-Jones parameters R_{ij}, A_{ij} from OPLSAA or from AutoDock and Hydrogen bond parameters $\tilde{R}_{ij}$, $\tilde{A}_{ij}$ from AutoDock.

3 Results

Screen with rigid (1e2n) receptor: A receptor has many possibilities to adapt to different inhibitors. Choosing one fixed receptor for all type of ligands restricts the amount of possible binding modes to only a few and therefore some ligands loose their specificity to this receptor. In 1e2n the receptor cavity is rather wide and many different ligands can fit into the cavity; but distinctive binding modes are missing for many ligands. Therefore the 10 known substrates of TK are energetically close to each other (see fig 1, left side), but because of the lack of specific binding modes they score worse than many ligands of the database.

Screen with flexible (1e2n) receptor: To model the receptor flexibility, we made 6 bonds of 4 side-chains flexible to allow the substrates to find their characteristic binding motif. Compared with the database all 10 ligands get lower affinities now (see fig 1, right side).

Astex data set results: The results from table 1 show that *FlexScreen* is either of similar accuracy (Glide) or significantly more accurate (Gold, FlexX). Additionally *FlexScreen* proved to reliably find the correct binding modes. In **89%** of the cases FlexScreen yielded a binding mode with a RMS deviation of less than 2.0 Å and performed therefore better than Gold, Glide and FlexX.

	FlexScreen	Glide	Gold	FlexX:
FlexScreen wins/total		33/59	20/27	40/59
Results < 2.0Å	77/86	41/59	20/27	32/59

Table 1. Summary of the docking results; compared are the RMS values of the different docking codes for the Astex data set and the percentage of cases having a RMS $<$2.0 Å.

4 Conclusion

Using side-chain flexibility (15 selected rotational bonds), all substrates ranked within the upper 10% of the database. The binding energy is substantially lowered for all of the ligands which supports the assumption that the receptor is now sufficiently able to adopt to the docking ligand and to model their specific binding motif, which can be compared with the x-ray receptor-ligand complex. Concerning the accuracy and reliability of finding the experimental binding mode *FlexScreen* proved to be of better performance than the three other docking programs (Glide, Gold and FlexX).

5 Acknowledgments

We thank the Fond der Chemischen Industrie, the BMBF, the Deutsche Forschungsgemeinschaft (grant WE) and the Kurt Eberhard Bode Stiftung for financial support.

References

1. W. Wenzel and K. Hamacher, *Stochastic Tunneling Approach for Global Minimization of Complex Potential Energy Landscapes*, Phys. Rev. Lett. **82**, 3003–3007, 1999.
2. H. Merlitz, B. Burghardt, and W. Wenzel, *Application of the Stochastic Tunneling Method to High Throughput Database Screening*, Chem. Phys. Lett. **370**, 68–73, 2003.
3. B. Fischer, H. Merlitz, and W. Wenzel, *Increasing Diversity in In-Silico Screening with Target Flexibility*, CompLife, 186–197, 2005.

Mutations as Trapdoors: The Rop-Dimer with Two Competing Native Conformations

Alexander Schug and José Onuchic

Center for Theoretical Biological Physics, University of California, San Diego, USA
E-mail: aschug@ctbp.ucsd.edu

Conformational transitions are the molecular mechanism for regulating protein function. Structure-based models are a computational tractable way to simulate these transitions. A model able to accommodate multiple folding basins is proposed to explore the mutational effects in the folding of the Rop-dimer (ROP). In experiments, ROP mutants show unusually strong increases in folding rates with marginal effects on stability. We investigate the possibility of two competing conformations representing a parallel (P) and the wild-type (WT) anti-parallel (AP) arrangement of the monomers as possible native conformations. We observe occupation of both distinct states and characterize the transition pathways. An interesting observation from the simulations is that, for equivalent energetic bias, the transition to the P basin (non WT basin) shows a lower free-energy barrier. Thus the rapid kinetics observed in experiments appears to be the result of two competing states with different kinetic behavior, triggered upon mutation by the opening of a trapdoor arising from the Rop-dimers symmetric structure. The general concept of having competing conformations for the native state goes beyond explaining ROP's mutational behaviors and can be applied to other systems. A switch between competing native structures might be triggered by external factors to allow, for example, allosteric control or signaling.

1 Introduction

The funneled energy-landscape and the principle of minimal frustration explain protein folding as a diffusive process. Multiple routes lead from the unfolded to the folded state.[1,2] Evolutionary pressure smoothened the underlying energy-landscape sufficiently that local minima or roughness do not interfere with folding. The resulting bias towards the native state is robust, so that changes in environmental conditions or limited mutations change neither the structure of a protein nor its folding behavior.

Structure-based simulations, based on the work of Go[3], use these ideas and stipulate that folding can be simulated *in-silico* by only taking native interactions into account[2,4]. Commonly, one coarse-grained the description of a protein in such simulations. Each amino-acid is described as a single bead centered on the position of the C_α-atoms.

2 Simulations on the Rop-dimer

ROP (repressor of primer) [a] is a homodimer of 2x63 amino acids(AA). It is part of a genetic control mechanism in the ColE1 plasmid system and binds to RNA. Its wild-type (WT) structure is that of a coiled-coil helix bundle with each monomer consisting of two helices (see Fig.). The two monomers are arranged anti-parallelly (AP) for the WT. In a series of experiments, the hydrophobic interface between the two monomers has been

[a] ROP is sometimes also called ROM (Regulator of RNA I).

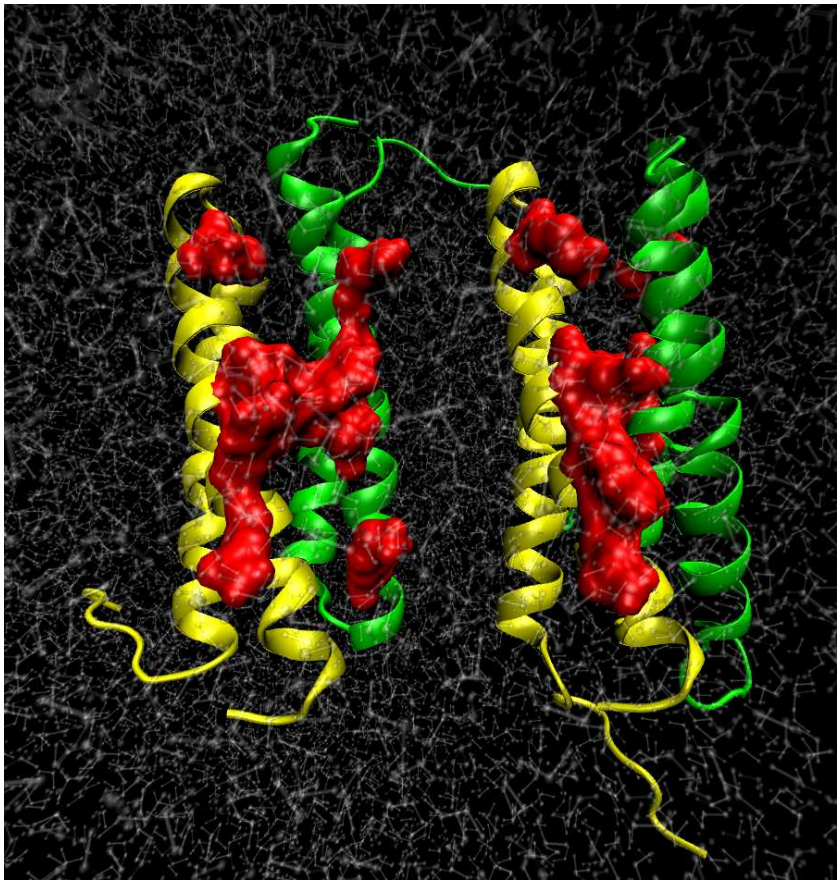


Figure 1. **Structures and symmetry of the Rop-dimer**
Hydrophobic core mutations have a strong effect on the folding/unfolding kinetics and RNA-binding ability of the Rop-homodimer. They can change the WT anti parallel arrangement (left) of monomers (in green and yellow) to a parallel one (right). This conformational transition disrupts the charged RNA-binding interface (highlighted red sidechains) and makes the parallel mutant dysfunctional. A dual-funneled energy landscape with two competing native structures explains Rop's mutational behavior.

mutated.[5,6] These mutations resulted in strong changes of up to four order of magnitude in the folding/unfolding rates. Specific mutants lost their RNA-binding ability, with partly discrepant behavior in vivo and in vitro.[7] One specific mutant has parallelly (P) arranged monomers and lost its RNA-binding ability. This is not surprising, as the ability to bind RNA is linked to an interface on the surface of the WT, which is disrupted in the P structure.

To better understand these experimental results, we assume that the mutations trigger the competition of P and AP by symmetrizing the interface[8,9]. We set up structure-based simulations with two competing native states[9] and observe transition between the unfolded (U), folded P and and folded AP states. It is important to note that we gave both P and AP an exactly equal energetic bias.

In the simulations we observe transitions from U to P and AP. These simulations allow

to derive a free-energy landscape. The free-energy barrier for the transition from U to P is roughly 20% smaller than the one from U to AP. Therefore the kinetics U-P are faster than for U-AP. This is a purely entropic effect resulting from the geometrically more accessible P conformation. In P the less mobile turns of the two monomers can face each other in the transition state ensemble, while the floppy tails of the monomers can still move freely. However in AP the turn face the floppy tail of the other monomer and are therefore more difficult to localize.[9]

ROP's accelerated kinetics for the mutants can therefore be understood as a result from the competition of P and AP. While in the WT the slow kinetics of AP dominate, the mutations in the hydrophobic core open the trapdoor to P and enable P as an off-pathway kinetic trap. One measures an increase in kinetics. [b]

It seems possible, that some mutants possess a degenerated native state, in which both P and AP are present. Especially the mutant $Ala_2Leu_2 - 6$ might express this behavior, as it both binds RNA but is also highly similar to the mutant $Ala_2Ile_2 - 6$, which only differs by possessing Ile instead of Leu for some AA in the hydrophobic core. Current experiments verify this prediction.

Mutant	Binds RNA *in*		Relative		Structure
	vitro	*vivo*	k_F	k_U	
WT	Y	Y	1	1	AP (X-Ray, NMR)
$Ala_2Leu_2 - 2$	Y	Y	3.2	18	AP (*in vitro* activity)
$Ala_2Leu_2 - 4$	Y	P	1.5	28	AP (*in vitro* activity)
$Ala_2Leu_2 - 6$	Y	N	310	31000	AP (*in vitro* activity)
$Ala_2Leu_2 - 8$	Y	N	610	50000	AP (*in vitro* activity)
$Ala_2Leu_2 - 6$-rev	Y	-	85	670	AP (*in vitro* activity)
$Ala_2Leu_2 - 8$-rev	Y	N	92	2700	AP (*in vitro* activity)
$Leu_2Ala_2 - (2+7)$	Y	-	10	18	AP (*in vitro* activity)
$Ala_2Ile_2 - 6$	N	N	-	-	P

Table 1. *Experimental data of the Rop-dimer and some mutants.*
The RNA-binding ability is present (Y), non-present (N) or partially present (P). The folding and unfolding rates k_F and k_U are given relative to the WT. The mutants are named according to the number and location of mutated hydrophobic core amino acid pairs.[5,6] The structures haven been determined only for the WT and the last mutant, all other structures have been assumed to be AP because of their RNA binding behavior.

Acknowledgments

This work was funded by the National Science Foundation-sponsored Center for Theoretical Biological Physics (Grants PHY-0216576 and 0225630) and also by Grant NSF MCB-0543906.

[b]The kinetics are determined experimentally by Circular Dichroism measurements which cannot distinguish between transitions from U to P or AP, as both states have equivalent helical content.

References

1. J. N. Onuchic and P. G. Wolynes, *Theory of protein folding*, Curr. Opin. Struct. Biol., **14**, 70–75, 2004.
2. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Funnels, pathways, and the energy landscape of protein folding - a synthesis*, Proteins, **21**, 167–195, 1995.
3. Y. Ueda, H. Taketomi, and N. Go, *Studies on Protein Folding, Unfolding, and Fluctuationsby Computer Simulation. II. A Three-Dimensional Lattice Model of Lysozyme*, Biopolymers, **17**, 1531–1548, 1978.
4. C. Clementi, H. Nymeyer, and J. N. Onuchic, *Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? an investigation for small globular proteins*, J. Mol. Biol., **298**, 937–953, 2000.
5. M. Munson, S. Balasubramanian, K. G. Fleming, A. D. Nagi, R. O'Brien, J. M. Sturtevant, and L. Regan, *What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties*, Protein Science, **5**, 1584–1593, 1996.
6. M. Munson, K. S. Anderson, and L. Regan, *Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude*, Fold. Des., pp. 77–87, 1997.
7. T. J. Magliery and L. Regan, *A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics*, Protein Engineering, Design & Selection, **17**, no. 1, 77–83, 2004.
8. Y. Levy, Sam S. Cho, T. Shen, J. N. Onuchic, and P. G. Wolynes, *Symmetry and frustration in protein energy landscape: A near degeneracy resolves the Rop dimer-folding mystery*, PNAS, **102**, no. 7, 2373–2378, 2005.
9. A. Schug, P. C. Whitford, Y. Levy, and J. N. Onuchic, *Mutations as trapdoors to two competing native conformations of the Rop-dimer*, PNAS, **104**, 17674–1767, 2007.

Global Motions in the Nucleosome Explored Using a Coarse-Grained Model

Karine Voltz[1,2], Joanna Trylska[3], Valentina Tozzini[4], Vandana Kurkal-Siebert[5], Jeremy C. Smith[2,6], and Joerg Langowski[1]

[1] Biophysics of Macromolecules, German Cancer Research Center, Heidelberg, Germany

[2] Computational Molecular Biophysics, IWR, University of Heidelberg, Germany
E-mail: Karine.Voltz@iwr.uni-heidelberg.de

[3] Interdisciplinary Center for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland

[4] NEST-INFM-Scuola Normale Superiore, Pisa, Italy

[5] Molecular Modelling, Polymer Physics, BASF Aktiengesellschaft, Ludwigshafen, Germany

[6] Center for Molecular Biophysics, Oak Ridge National Laboratory, USA

The nucleosome is the basic compacting unit of chromatin, a complex structure that enables DNA to fit in the eucaryotic cell nucleus. Because many biological processes require free DNA as a substrate, the nucleosome has to undergo conformational transitions to allow the DNA target sites to be exposed. To obtain insight into the global dynamics of the nucleosome, microsecond timescale coarse-grained molecular dynamics is performed. Here we report a principal component analysis (PCA) realised on a 5μs coarse-grained molecular dynamics trajectory to identify the global motions obtained in the nucleosome.

1 Introduction

In the nucleosome, 147-bp DNA are wrapped almost twice around a protein core consisting of eight histone proteins, one tetramer H3-H4 and two dimers H2A-H2B. The histone protein core, composed mainly of α-helices, is very stable in contrast to the non-structured histone N-terminal tails, that pass between the DNA superhelix turns. Despite the high number of interactions between nucleosomal DNA and the surface of the protein core, nucleosomal DNA can become free to be processed by DNA binding proteins involved in DNA transcription, replication, recombination or repair. The mechanisms underlying DNA target site exposure are still under debate. One proposed mechanism, based on single-nucleosome FRET evidence, postulates that significant pieces of nucleosomal DNA (up to 70 bp) can transiently detach from the protein core[1,2]. These reversible conformational changes, believed to occur on the 50-250 ms timescale, could be responsible for the progressive invasion of the nucleosome by DNA-binding proteins.

To characterize the conformational transitions leading to DNA accessibility, it is important to understand nucleosome dynamics at equilibrium and on a long timescale. For this purpose, we performed coarse-grained (CG) molecular dynamics (MD) simulations using a model specifically developed for the nucleosome[3]. Classical all-atom MD of large systems such as the nucleosome is currently limited to about 100 nanosecond timescale, while with the present model of the nucleosome, a one-bead representation together with the absence of explicit water, MD simulations of the nucleosome can be carried out over

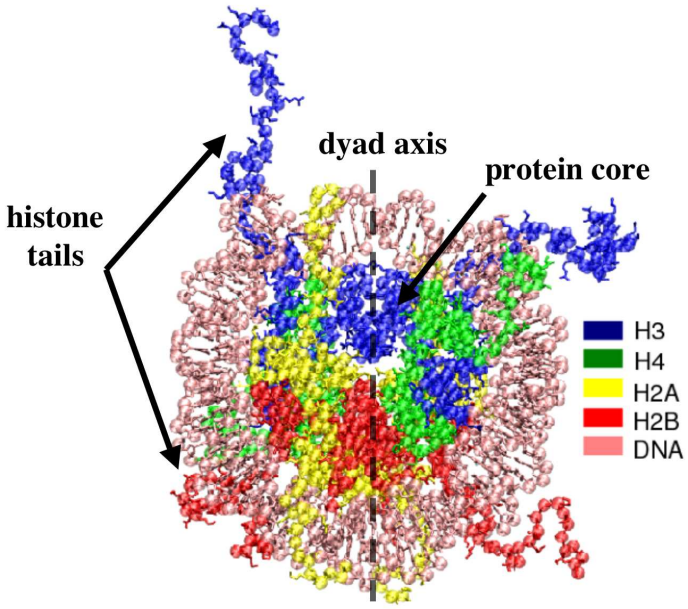


Figure 1. All-atom and coarse-grained representation of the nucleosome (structure 1KX5). The residues are represented by single spherical beads centered on α-carbon for amino acids and on phosphorus for nucleic acids. The histones and the DNA are represented in different colors. The nucleosome is symmetric with respect to the so-called dyad axis.

several microseconds. In the present proceedings, after briefly presenting the CG model, we report a principal component analysis (PCA) performed on a 5μs CG MD trajectory enabling the identification of slow collective motions in the nucleosome.

2 Method

A detailed description of the CG model and its force-field parameterization is reported in a previous study[3]. In this CG model, protein residues and DNA nucleotides are represented as single beads (Figure 1) interacting through harmonic (for neighboring) or Morse (for nonbonded) potentials which depend on the interbead distances. This model shares similarity with Gaussian network models (GNM) but, in contrast to GNM, the Morse description of nonbonded interactions allows realistic anharmonic dynamics of the system. Force-field parameters were estimated by Boltzmann inversion of the corresponding radial distribution functions computed from a reference 5-ns all-atom MD simulation and further refined to obtain agreement with all-atom MD root-mean-square fluctuations (RMSF).

Coarse-grained MD simulations were performed using the DL_POLY package[5]. The starting structure used was the 1KX5 structure[4] energy-minimized in solvent using the

Charmm[6]27 all-atom force-field. A Langevin bath was used to account for the frictional and stochastic effects of the solvent. The reduction of the system allowed us to apply a 20-fs timestep. The simulation was carried out at 300 K during 5 μs.

Principal component analysis, also called essential dynamics analysis[7], provides a way to identify the most significant directions of motions in the system. The method consists in diagonalizing the symmetric 3N$\times$3N covariance matrix derived from the MD trajectory, whose elements are defined as $C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$ where x_i and x_j are the Cartesian coordinates of the atom i and j. The resulting eigenvectors give the direction and their corresponding eigenvalues quantify the magnitude of the fluctuations. Eigenvectors with the highest eigenvalues are called the principal components or principal modes. Here, PCA was performed with the GROMACS package on a 5-μ CG MD trajectory. The tails were excluded from the analysis since their high mobility masked the motion of the DNA and the protein core.

2.1 Results and Discussion

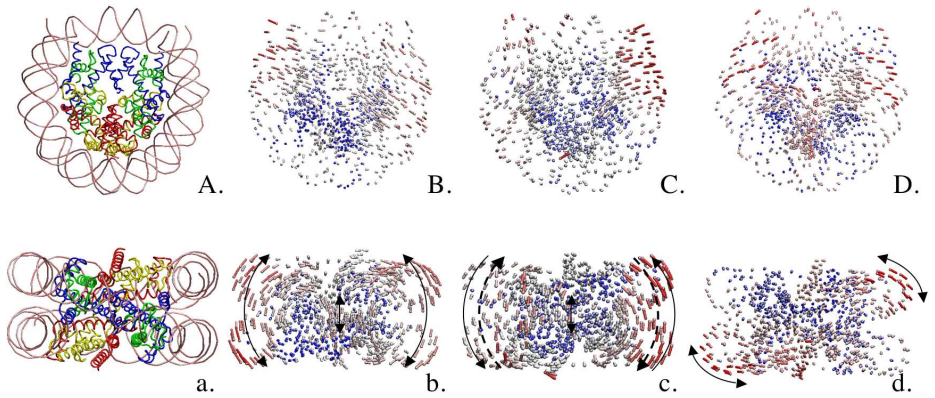


Figure 2. Global motions of the nucleosome. A and a: initial nucleosome 1KX5 structure. B–D and b–d: nucleosome conformations deformed along the first three modes. Each residue is colored with respect to its root-mean-squared fluctuation (RMSF) along the mode; blue and red correspond to low and high RMSFs, respectively.

Figure 2 shows the directions of the motions for the first three principal components. Residues are colored with respect to their RMSF along each eigenvector. The ten eigenvectors with the largest eigenvalues describe 22 % of the total protein motion, a smaller fraction than what is usually observed for other systems such as proteins. This result may be the consequence that important structural transitions in the nucleosome occur on a larger timescale than 5μs.

The motion along the first mode corresponds to a bending out of the plane of the nucleosome that mostly involves i) both extremities (ten basepairs) of the DNA superhelix; ii) the facing DNA stretches located on the respective opposite turn; and iii) amino acids of H3, H3', H4 and H4' interacting with these regions. Simultaneously to the bending of this

region, several residues centered around the dyad axis (i.e., ≈10-bp of DNA and residues from H3 and H3') translate in the opposite direction to the out-of-plane bending motion. The most rigid regions along this mode are residues that belong to the histones H2B and the DNA basepairs interacting with them.

The second principal component describes a rocking motion center of which is located on the dyad axis. As in the first mode, this mostly involves both extremities of the DNA superhelix, the facing DNA stretches located on the respective opposite turn, and amino acids interacting with these regions, while histones H2B and interacting DNA residues contribute less to this motion.

Finally, the third principal mode, represents a deformation of the nucleosome in the plane of the nucleosome. In this motion, the DNA superhelix extremities and amino acids interacting with these regions participate in the stretching of the nucleosome in a direction perpendicular to the dyad axis while H2B residues and the DNA part interacting with them contributes to stretching the nucleosome in the direction of the dyad axis.

3 Outlook

Further work is ongoing to investigate, in particular, the spontaneous unwrapping of nucleosomal DNA extremities from the nucleosome protein core. We also plan to study the interactions between several nucleosomes in a nucleosomal array within the context of the chromatin fiber.

References

1. G. Li, M. Levitus, C. Bustamante and J. Widom, *Rapid spontaneous accessibility of nucleosomal DNA* Nat. Struc. Mol. Biol. **12**, 46–53, 2005.
2. M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova, S. H. and Leuba, *Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer* Proc. Natl. Acad. Sci. U.S.A. **102**, 3278–3283, 2005.
3. K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Langowski and J. Smith, *Coarse-grained force field for the nucleosome from self-consistent multiscaling* J. Comput. Chem. **29**, 1429–1439, 2008.
4. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9Å* J. Mol. Biol. **319**, 1097–1113, 2002.
5. W. Smith, T. R. Forester, *DLPOLY2.0: A general-purpose parallel molecular dynamics simulation package* J. Mol. Graph. **14**, 136–141, 1996.
6. Jr. A.D. MacKerell, N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids* Biopolymers **56**, 257–265, 2001.
7. A. Amadei, A. B. M. Linssen, H. J. C. Berendsen *Essential dynamics of proteins* Proteins **17**, 412-425, 1993.

Effects of Confinement and Crowding in Protein Folding

Michal Wojciechowski and Marek Cieplak

Institute of Physics, Polish Academy of Sciences,
Al. Lotników 32/46, 02-668 Warsaw, Poland
E-mail: mwoj@ifpan.edu.pl

We consider folding and unfolding of a protein contained within a sphere of radius R. We use a coarse-grained geometry based model. We find that the folding time is essentially independent of R but the possible unfolded structures do depend on R. Several proteins placed within the sphere fold mostly independent of one another unless one introduces attractive interactions between them. Attractive interactions between molecules have strong influence on the folding times.

1 Introduction

Many experimental and theoretical studies of proteins involve very low protein concentrations. In cells, however, the proteins are confined to compartments and come with high concentrations[1]. It is thus interesting to assess the role of confinement and crowding theoretically by first considering simple models.

Previous theoretical studies of confinement[2–6] have been focused on thermodynamics. The confinment has been found to lead to a greater thermodynamic stability, broader and taller specific heat and more compact unfolded conformations[4,3].

Recently, we have studied effects of molecular crowding and caging on protein folding within a simple molecular dynamics model[7]. Here, we provide a brief account of these studies.

2 Model

We used the coarse-grained Go-like model, where amino acids are represented by beads. The beads interact by a potential that maintains chain topology and enforces the local backbone stiffenss through the chirality terms[8]. The native contacts are described by

$$V_{ij}^{contact} = 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1}$$

where r_{ij} describes the distance between the bead centers. By taking $\sigma_{ij} = r_{ij}^{native}/\sqrt[6]{2}$ the potential acquires a minimum at a distance r_{ij}^{native} that is found in the native structure of the protein. The repulsive potential that keeps protein inside the sphere may be written as

$$V_{wall}(s_i) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{s_i} \right)^{12} - \left(\frac{\sigma}{s_i} \right)^{6} \right] + \epsilon \ , & s_i < 4 \qquad s_i = R - r_i \\ 0 \qquad\qquad\qquad\qquad\qquad\quad , & s_i \geq 4 \qquad \sigma = 4/\sqrt[6]{2} \end{cases} \tag{2}$$

where r_i is a distance of the ith center to the origin of the sphere, R is the radius of the sphere.

The attractive interaction between separate molecules is introduced in simplified way. When dealing with several identical proteins, it is natural to assume that if a pair of amino acids forms the native contact in one molecule, the same pair of amino acids, but belonging to different molecules, will also interact. The strength, ϵ_I, of such an interaction is an input to the modelling. We take

$$V_{ij'}^{inter} = 4\epsilon_I \left[\left(\frac{\sigma_{ij}}{r_{ij'}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij'}} \right)^{6} \right], \qquad 0 \geq \epsilon_I \geq \epsilon \tag{3}$$

and ϵ_I is considered to be weaker than ϵ from Eq.1. We have determined the folding times, t_{fold}, defined as the median over the set of 301 trajectories. The folding time of a molecule is defined as the first time at which all of its native contacts are established simultaneously. For several molecules, test of the presence of the native contacts are performed for each molecule separately. The simulations were performed for crambin – the small, one domain, α-β protein with the PDB code 1CRN.

3 Results

We computed t_{fold} of crambin in spheres with various radii ranging from $R = 1000$Å down to 18Å (corresponding to the smalest sphere that may fit the crambin in the native state). The starting structures for folding were obtained through thermal unfolding at a high temperature in sphere of radius R_0. To evaluate the influence of the starting structures on t_{fold} we also study folding in an unbounded space. Results of the simulations are presented on Fig.1A.

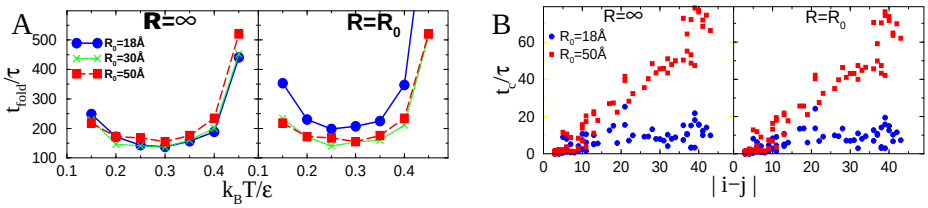


Figure 1. A: Folding time of crambin computed with different starting structure sets (corresponding to various indicated choices of R_0). The left panel shows folding in an infinite space and the right panel in a sphere of a finite radius radius $R = R_0$. B: Folding scenarios of crambin computed with different starting structure set (R_0 describes the radius of sphere in the unfolding simulation that produce the starting structures for folding). Left panel shows folding in infinite space, right panel folding in sphere of radius $R = R_0$ (the same radius as for the unfolding procedure).

There is no influence of the starting structures on t_{fold}, Fig.1A (left panel), and there is only a weak influence the sphere size on t_{fold} as shown in Fig.1A (right panel). Thus the confinement does not change the folding times of proteins unless one considers very tight confinement conditions under which the protein barely moves. We also find that even

though the sequencing of folding events depends on the choice of the starting R_0, the confinement itself does not affect it, see Fig1B.

We now consider the crowding phenomena and place several protein in one 36Å-sphere. We calculate t_{fold} for one, two, eight and twelve molecules. The simplest case is when the only inter-molecular interactions are only due to the excluded volume. The corresponding results are presented in Fig.2A. It is seen that the number of molecules that are present does not change an individual folding time.

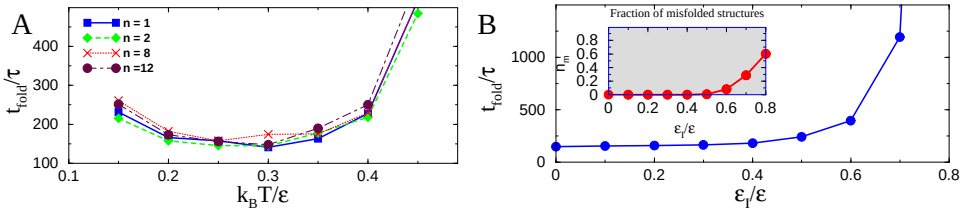


Figure 2. A: Folding times for different number of proteins within the sphere. B: Folding times for four crambins on different interaction strength. The inset shows the fraction of misfolded structures.

The situation changes when attractive inter-molecular interactions are added. Fig.2B shows t_{fold} for R=36 Å. For $\epsilon_I < 0.5\epsilon$, t_{fold} is nearly the same as for ϵ_I=0. However, for larger values of ϵ_I, t_{fold} raises significantly. Eventually, all structures get misfolded, because of aggregation.

Acknowledgments

This work has been supported by the grant N N202 0852 33 from the Ministry of Science and Higher Education in Poland.

References

1. R. J. Ellis and A. P. Minton, *Protein aggregation in crowded environments*, Biol Chem. **387**, 485–97, 2006.
2. A. Baukmketner, A. I. Jewett, and J. E. Shea, *Effect of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape.*, J. Mol. Biol. **332**, 701–13, 2003.
3. N. Rathore, T. A. Knotts IV, and J. J. de Pablo, *Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations*, Biophys J. **90**, 1767–73, 2006.
4. F. Takagi, N. Koga, and S. Takada, *How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations*, Proc. Natl. Acad. Sci. USA **100**, 11367–72, 2003.
5. A. I. Jewett, A. Baumketner, and J.-E. Shea, *Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway*, Proc. Natl. Acad. Sci. USA **101**, 13192–7, 2004.

6. M. S. Cheung, D. Klimov, and D. Thirumalai, *Molecular crowding enhances native state stability and refolding rates of globular proteins*, Proc. Natl. Acad. Sci. USA **102**, 4753–8, 2005.
7. M. Wojciechowski and M. Cieplak, *Effects of confinement and crowding on folding of model proteins*, Biosystems (in press), 2008.
8. J. I. Kwiecinska and M. Cieplak, *Chirality and protein folding*, J. Phys. Cond. Mat. **17**, S1565–80, 2005.

Conformational Study of Alzheimer's Aβ Wild Type Peptide and Flemish Mutant

Priya Anand[1,2], **F. S. Nandel**[1], **and Ulrich H. E. Hansmann**[2]

[1] Department of Biophysics, Panjab University, Chandigarh-160014, India

[2] John von Neumann Institute for Computing, Research Centre Jülich, 52425 Jülich, Germany
E-mail: priyaanand27@gmail.com

Using replica exchange Molecular Dynamics simulations of Aβ peptide, we compare the ensembles of monomer configurations of the Flemish (A21G) familial Alzheimer's disease mutant the wild type (WT) sequence. Our simulations start from random conformations. We find that similarities between the Flemish and WT Aβ peptide in terms of the random coil like structure. At room temperature in implicit solvent, the Aβ_{1-39} monomer does not adopt a unique folded conformation but adopts one of the several low energy structures, with the U-shape configurations that are strongly amphipathic. No beta content is observed which therefore seems to be a product of oligomerization and aggregation.

1 Introduction

Alzheimer's disease (AD) is a neurological disorder, affecting approximately 12.5 and 47.2% of the population in the United States over the ages 65 and 85, respectively. Formation of amyloid fibrils is the hallmark of Alzheimer's disease. These Aβ peptides are released from proteolytic cleavage of the amyloid precursor protein (APP) as Aβ39 or Aβ42 residue sequence with unknown function. Many familial Alzheimer's disease mutants of the APP protein are external to the Aβ peptide sequence, and thus influence Aβ processing, but some set of mutations which cluster near amino acid positions 21 through 23 in the amyloid β peptide possibily changes peptide biochemistry[1]. Some well studied point mutations are Flemish (A21G), Arctic(E22G), Italian (E22K), Dutch(E22Q), Iowa(D23N) and double Dutch/Iowa mutants [2]. Despite point mutations near 21/22 region, these mutations show strong differences in the kinetics of the formation of fibril assemblies as compared to wild type Aβ peptide and in vitro studies have found that A21G (Flemish) mutations slower aggregation kinetic as compared to WT Aβ and E22G (Arctic) mutations. Hence, this distinction is important for understanding the mechanism of amyloid assembly and critical for correctly assigning the identity of NMR cross peak resonances in 2D structural analysis of Aβ, where it has been assumed that the peptide is monomer. Thus the present study focus on the effect of mutation (A21G) on the monomer structure of Aβ39 Wild Type(WT) by performing molecular dynamics simulations.

2 Material and Methods

The peptide Aβ originally consists of 42 amino-acid residues: [Asp-Ala-Glu-Phe-Arg-His-Asp-Ser-Gly-Tyr-Glu-Val-His-His-Gln-Lys-Leu-Val-Phe-Phe-Ala-Glu-Asp-Val-Gly-Ser-Asn-Lys-Gly-Ala-Ile-Ile-Gly-Leu-Met-Val-Gly-Gly-Val-Val-Ala-Ala], which is usually expressed as Aβ. Replica exchange Molecular Dynamics (MD) and canonical MD

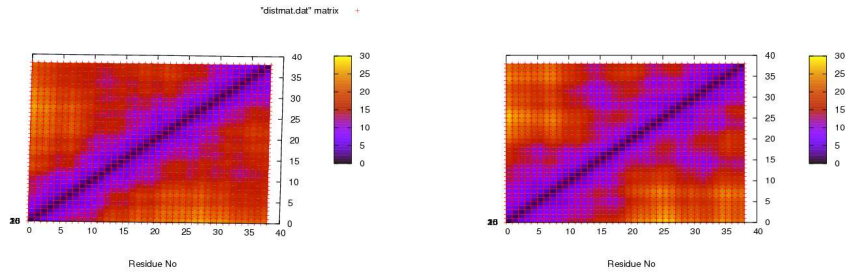


Figure 1. Contact map dor Cα-Cα for WT and Flemish mutant.

simulations are carried out with the AMBER9 program, using all atom force field ff99. The effect of solvation is approximated by a generalized Born solvent implicit solvent model, with random starting geometry. Sixteen replicas are simulated at temperatures range of 200 K to 640 K and periodically swapped between neighboring temperatures. Exchange attempts are made after every 0.02ps of simulation. High temperature simulation segments facilitate the crossing of the energy barriers while the low temperature ones explore in detail energy minima[3,4]. The Shake algorithm is used to constraint all bond lengths. In the canonical simulation, temperature is set by velocity reassignment from a Maxwell-Boltzmann distribution at 291 K and maintained at that temperature by using a Langevin thermostat. About 5,000 steps of minimization is followed by an initial equilibration run.

3 Results and Discussion

In order to probe the structure of Aβ_{39} monomer we have performed multiple replica-exchange and regular canonical molecular dynamics (at T=291K) simulations starting with different initial configuration of flemish mutant and WT Aβ peptide. Replica exchange molecular dynamics simulations results indicate that each replicas perform a random walk in temperatures ladder, allowing the replicas to escape local minima. As a consequence, reliable physical quantities are calculated over the whole range of temperatures.

At room temperature, the WT Aβ monomer does not have a unique folded conformation but adopts one of the several low energy structures. The root mean square displacement based clustering analysis [5] shows that the ensemble at $291K$ consists of three clusters differ little in their average potential energies (cluster A: 946.09(4) kcal/mol, cluster B: 962.59(2) kcal/mol,cluster C: 949.10 (2) kcal/mol, data for WT Aβ peptide)[6]. All three cluster also share as a common theme the U-shape of the configurations that are strongly amphipathic. The longer arm is made mostly from hydrophilic and charged amino acid residues while the shorter arm is formed exclusively hydrophobic branched residues. However, no beta content is observed in both WT and flemish mutatnt(1.4%), which therefore seems to be a product of oligomerization and aggregation. A more thorough analysis reveals differences between the various clusters. Cluster A, is a random coil structure with turn around Ala21-Asn27 and appears with a frequency of 58 % in WT Aβ peptide and

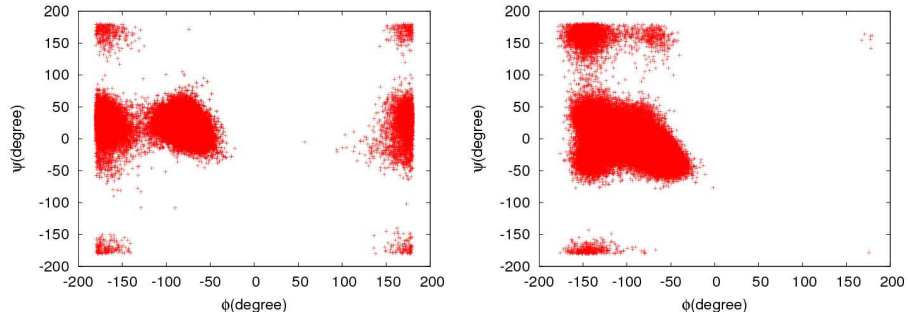


Figure 2. Ramachandran Plot for residue Ala21 and Gly21 in WT and flemish mutant.

30% in Fleinsh mutant. We observe that the turn around residues 21-28 is conserved, however having lower probability in flemish mutant as compared to WT, with turn stabilized by hydrophobic interactions between Val24 and Asn27,Gly25 and Lys28 and electrostatic interactions between Lys28 and Glu22 in WT peptide. The region Leu17-Ala21 in WT Aβ has a well defined structure and shows considerably smaller structural fluctuations than rest of the peptide, which is consistent with NMR studies[7,8] that suggest a defined structure around central hydrophobic region and indicates that this region is critical for fibril formation. Configurations of cluster B also shows a random coil like structure with a turn around Val12-Leu17. This cluster appears with a frequency of 29% in WT and 31% in flemish mutant. In these configurations we observe in WT the salt bridges between residue Asp23-Lys28/16 that have been reported in various experiments as controlling the aggregation rate of Aβ[9]. Finally, 13% of configurations are rich in helicity and form cluster C present only in WT Aβ peptide. Configurations in this cluster are similar to the structure that was determined in a 40% TFE/water solution.[10].

In femish mutant, Ala21Gly we observe another cluster populated with 39 % that has turn shifted from Glu22-Lys28 region to 16-23 region. Gly at position 21 might diminish the N-terminal $beta$ strand integrity in amyloid fibril by shifting the turn region, resulting into qualitatively different behavior in the structural integrity of the protofibril. In order to further quantify the variation of the turn region, we calculated the intramolecular distance between N atom of the Lys28 and carbony oxygen of Glu22 (distance around 10Å) and various other hydrophobic interaction between Val24 and Asn27, Gly25 and Lys28 (distance around 8Å) present in WT Aβ peptide, as expected the interactions stabilizing the turn region around 21-28 are not present in this cluster.

Fig. 1 shows the intramolecular Cα-Cα contact map for Flemish mutant, indicates more populated interactions around region 16-23 as compared to WT Aβ peptide. However the RMSD plot of the Cβ Aβ_{30} WT and Flemish mutant displays that similar trend with radius of gyration around 10Åfor both the peptides, indicating similar size or end to end distance Cα distances for both the monomer.The Ramachandran plot of residues Ala21 in WT and Gly21 in flemish mutant (Fig. 2) also explores essentially the same the favored regions of ϕ,ψ, indicating no much effect on the flexibility of the monomer around this region. The mean percentage of α-helix, turn and β contents for all the residues in both Aβ peptide, percentage of beta strand is 1.2 % and 35-40% for α helix in flemish mutant, however no

change in the secondary structure content observed around resiude Ala21 and Gly21 in both WT and flemish mutant.

We found that the region around 17-21 which is conserved in WT Aβ was not well conserved in flemish mutant. Even the new cluster lack the native interactions present in WT Aβ peptide around the turn region, with some new interactions not specific to native. In native Aβ peptide the lactam bridge between Asp23 and Lys 28 increases the fibrillogenesis rate by three orders of magnitude, however this bridge is disrupted in flemish mutant which might explain the possible role of this mutant on slowing the aggregation kinetics of Aβ peptide and further stabilizing the monomer structure.

Acknowledgments

This work is supported by DAAD & Forschungszentrum Jülich grants. All calculations are done on Nicole Cluster of NIC, Research Centre Jülich, Germany.

References

1. Goedert, M. and M.G. Spillantini. Science **134,**777-781 (2006).
2. Christian Haass and Harald Steiner Nature Neuroscience **9,** 859-860 (2001).
3. U. H. E. Hansmann, Chem. Phys. Lett. **281,** 140-150 (1997).
4. Y. Sugita,and Y. Okamoto, Chem. Phys. Lett.**314,** 141-151 (1999).
5. Michael Feig, John Karanicolas, Charles L. Brooks III, Journal of Molecular Graphics and Modeling **22,** 377-395 (2004) .
6. Priya Anand, F.S. Nandel and U.H.E. Hansmann J. Chem. Phys. **128,** 165102-06 (2008) .
7. R. Riek, P. Guntert, H. Dobeli, B. Wipf, and Wu Trich, K., Eur. J. Biochem. **268,** 5930-5936 (2001).
8. S. Zhang, K. Iwata, M. J. Lachenmann, J. W. Peng, S. Li, E. R. Stimson, Y. A. Lu, A. M. Felix, J. E. Maggio, and J. P. Lee, J. Struct. Biol. **130,** 130-141 (2000).
9. K. L. Sciarretta , D. L. Gordan, A. T. Petkova, R. Tycko, and S. C. Meredith, Biochemistry, **44,** 6003-6014 (2005).
10. P. E. Fraser, J. T. Nguyen, W. K. Surewicz, and D. A. Kirschner, Biophys. J. **60,** 1190-1201 (1991).

Effect of Temperature on the Structural and Hydrational Properties of Human Islet Amyloid Polypeptide in Water

Maximilian N. Andrews, Ivan Brovchenko, and Roland Winter

Department of Chemistry, Biophysical Chemistry, Dortmund University of Technology, D-44227 Dortmund, Germany
E-mail: {*maximilian.andrews, ivan.brovchenko, roland.winter*}*@tu-dortmund.de*

Structural and hydrational properties of the full-length human islet amyloid polypeptide 1-37 (hIAPP) were studied in a temperature range from 250 to 450 K by molecular dynamics computer simulations. At all temperatures studied, hIAPP does not adopt a well-defined conformation. The distribution of residues having the dihedral angles ϕ and ψ within the allowed regions of the Ramachandran plot which define β-sheets and poly(L-proline) II structures along the peptide chain is close to random, whereas a clear trend towards cooperative "condensation" is seen for residues having Ramachandran angles which characterize α-helices. This cooperativity and the number of intrapeptide H-Bonds is suppressed by heating or by introducing the natural intramolecular disulfide bond between residues 2 and 7. Intrinsic volumetric properties of hIAPP were estimated by taking into account the difference in the volumetric properties of hydration and bulk water. The temperature dependence of the density of hydration water indicates that the effective hydrophobicity of the hIAPP surface is close to that of carbon-like surfaces. The thermal expansion coefficient of hIAPP is found to be negative and decreases continuously upon heating from $\sim -3 \cdot 10^{-4}$ to $\sim -2 \cdot 10^{-3}$ K^{-1}. The spanning H-bonded network of hydration water at the hIAPP surface breaks via a percolation transition at about 320 K, which may be related to the drastic speed up of hIAPP aggregation seen experimentally in this temperature region.

1 Introduction

The aggregation of the human islet amyloid polypeptide (hIAPP) is involved in Diabetes Mellitus Type II. Hence, knowledge of the conformational behavior of this peptide is important for understanding the aggregation mechanism of hIAPP and for finding the means to prevent formation of its ordered fibrillar aggregates, which may be the main cause of decease. Experimental studies of the structural properties of hIAPP have not been successful due to its strong propensity to aggregate.

2 Systems and Methods

In this work, we performed MD computer simulation studies of the structural and hydrational properties of a single hIAPP peptide in liquid water in the temperature range from 250 to 450 K. All atomic molecular dynamics simulations were carried out with GROMACS v.3.3.1 using the OPLS-AA/L force field for the peptide and SPCE water molecules. Initially, the peptide was prepared in various starting conformations, including an α-helical conformation, four random conformations obtained from 1 ns runs at 1000 K *in vacuo*, of which one of the initial conformations being a fully extended isolated β-strand. After 15 to 30 ns simulation runs in water, the conformational behavior of hIAPP no longer depended on the initial configuration used. After 50 ns of equilibration at each temperature studied,

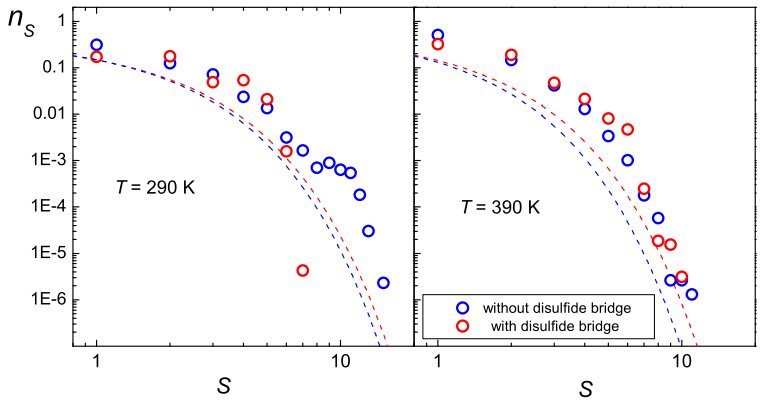


Figure 1. Probability n_S to find S successive residues with helical conformation. The dashed lines show n_S for a random distribution of residues in an infinite chain with the same content p of residues with analogous structure: $n_S = (1 - p)^2 p^S$.

200 ns trajectories were used for the analysis of the system properties. Two moieties of hIAPP were studied: hIAPP with and without the natural disulfide bridge between C2 and C7 residues.

3 Structural Properties

Analysis of the secondary structure shows that at all the temperatures studied, hIAPP does not adopt a well-defined conformation. The helical content of hIAPP, estimated as a fraction of residues having the dihedral angles within the allowed region of the Ramachandran plot, do not depend noticeably on the presence of a disulfide bridge and decrease upon heating. However, the ability of the helical residues to form a continuous sequence along the peptide chain is strongly suppressed by the disulfide bridge. This can be seen from the comparison of the probability distributions n_S to find S successive residues with helical conformation shown in Fig. 1. Large clusters of residues with helical dihedral angles disappear by introducing the disulfide bridge and by heating.

4 Volumetric Properties

The intrinsic volumetric properties of a biomolecule in water can be studied, when the density of hydration water is known.[1] The temperature dependence of the density ρ_h of the hydration water in a shell 0.3 nm thick at the hIAPP surface and of the density ρ_b of a bulk liquid water are shown in Fig. 2: ρ_h is below ρ_b and its temperature dependence is essentially linear. The temperature dependence of the density of hydration water indicates that the effective hydrophobicity of the hIAPP surface is close to that of carbon-like surfaces. Knowing the temperature dependences of ρ_h and ρ_b, we can estimate the intrinsic volume V_{int} of hIAPP from the equation: $V_{int} = V_{app} - \Delta V$. Here, V_{app} is the apparent volume of hIAPP measured as the difference between the volumes of the simulation boxes

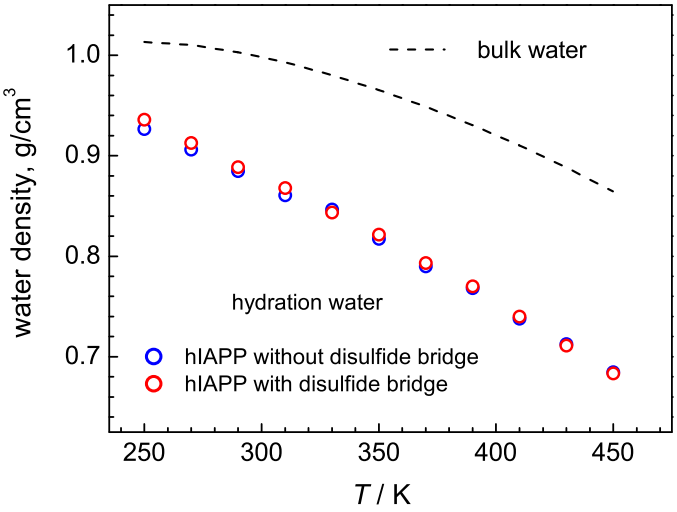


Figure 2. Temperature dependence of the density of bulk water and hydration water near hIAPP.

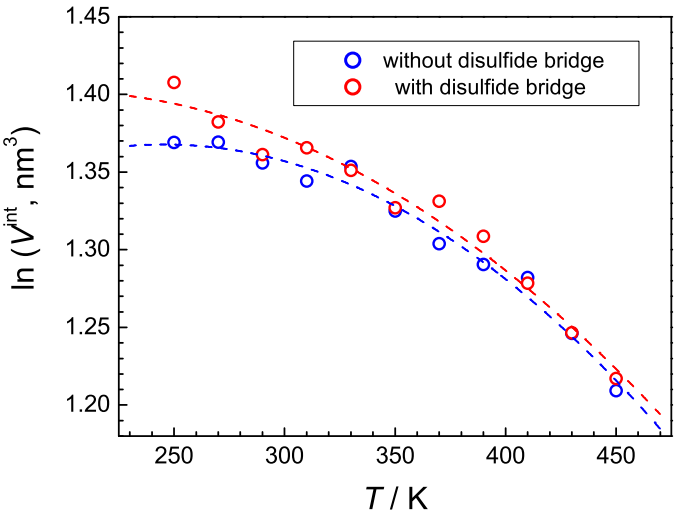


Figure 3. The temperature dependence of the logarithm of V_{int} of hIAPP. The lines are the fits to a quadratic polynomial.

with and without hIAPP, respectively, both having the same number of water molecules. The term ΔV accounts for the change of the system volume due to the different densities of hydration and bulk water, $\Delta V = V_h(1 - \rho_h/\rho_b)$, where V_h is the volume of hydration water. In a first approximation, V_h is the product of the solvent accessible area and the thickness of the hydration shell. The temperature dependence of the logarithm of V_{int} of

hIAPP is shown in Fig. 3. The slope of this dependence is equal to the intrinsic thermal expansion coefficient α_{int}. Similarly to the case of amyloid β peptide (1-42)[1], α_{int} of hIAPP is negative and becomes even more negative upon heating. Such behavior can be attributed to a decreasing helical content and a decreasing number of intrapeptide H-bonds. Note, that the disintegration of large clusters of helical residues by the disulfide bridge at low temperature (see Fig. 1) makes α_{int} more negative (see Fig. 3).

5 Thermal Disruption of the Hydration Water Network at the hIAPP Surface

The spanning H-bonded network of hydration water, which covers hIAPP homogeneously at low temperatures, breaks via a quasi-2D percolation transition, whose midpoint is located at about 320 K. Interestingly, approximately at this temperature, the experimentally measured lag time of hIAPP aggregation drops drastically[2]. Hence, we might conclude that the breakdown of the spanning H-Bonding network of hydration water might foster hIAPP aggregation.

Acknowledgments

Financial support from the International Max-Planck Research School in Chemical Biology and from the Federal State of NRW and the EU (Europäischer Fonds für regionale Entwicklung) is gratefully acknowledged.

References

1. I. Brovchenko, R. R. Burri, A. Krukau, A. Oleinikova, and R. Winter, *Intrinsic thermal expansivity and hydrational properties of amyloid peptide Aβ42 in liquid water*, to be published.
2. R. Kayed *et al.*, *Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation in vitro*, J. Mol. Biol., **287**, 781, 1999.

A Knowledge-Based Potential for Protein-RNA Docking

Ranjit P. Bahadur and Martin Zacharias

School of Engineering and Science, Jacobs University Bremen,
Campus Ring 1, D-28759, Bremen, Germany
E-mail: {m.zacharias, r.bahadur}@jacobs-university.de

Protein-RNA interactions play an important role in all cellular processes and it is important to understand the driving forces that govern this interaction. The mechanism by which a protein molecule specifically recognizes a RNA molecule in the cellular environment is not completely known. Here we developed a pair-potential function from the analysis of the 81 non-redundant atomic structures of protein-RNA complexes taken from the Protein Data Bank. This function helps us to understand the specificity of the interactions and could be useful in a protein-RNA docking algorithm where one tries to predict the correct complex structure starting from the individual components.

1 Introduction

Protein and RNA often interact in the cellular environment to perform essential cellular functions such as expression of a gene and its regulation. They can form binary complexes, for example, the aminoacyl-tRNA synthetases bind specific tRNAs for the translation of the genetic code; or multiple RNA and protein molecules can build a complicated cellular machine like a ribosome used for protein synthesis. To understand the functional mechanism of these complexes we have to elucidate the specificity of their interactions.

Several studies have been carried out recently to understand the structural basis of protein-RNA recognition [1-5]. All these methods consider the detailed atomic structures of the biomolecules. Here, we present an alternative approach to represent the protein and RNA chains in a reduced coarse-grained model, where each amino acid is represented by up to four pseudo atoms and each nucleotide by up to five pseudo atoms. We have calculated the pairwise contacts between the pseudo atoms of polypeptide and nucleotide chains and used them to derive a knowledge-based potential from a non-redundant dataset of 81 protein-RNA complexes recently compiled by Bahadur et al. [5]. Furthermore, the potential was included in a protein-RNA docking algorithm which can be used to predict complex structures starting from the individual structures of protein and RNA.

2 Materials and Methods

The dataset consist of 81 non-redundant known protein-RNA complexes taken from the PDB [6]. We have first translated the protein and RNA subunits into a reduced pseudo atom model. In case of the protein the same representation as implemented in the Attract docking program [7] was used. Briefly, each amino acid residues are represented by up to four pseudo atoms, two for main chain (N and O) and two for side chains. The side chains of Ala, Asn, Asp, Cys, Ile, Leu, Pro, Ser, Thr and Val are represented by one pseudo atom located at the center of geometry of all side-chain heavy atoms. Other larger side chains are represented by two pseudo atoms. The main chains pseudo atoms for all residues are rep-

Amino acids	Pseudo atoms
Ala	N, O, SC1
Arg	N, O, SC1, SC2
Asn	N, O, SC1
Asp	N, O, SC1
Cys	N, O, SC1
Gln	N, O, SC1, SC2
Glu	N, O, SC1, SC2
Gly	N, O, CA
His	N, O, SC1, SC2
Ile	N, O, SC1
Leu	N, O, SC1
Lys	N, O, SC1, SC2
Met	N, O, SC1, SC2
Phe	N, O, SC1, SC2
Pro	N, O, SC1
Ser	N, O, SC1
Thr	N, O, SC1
Trp	N, O, SC1, SC2
Tyr	N, O, SC1, SC2
Val	N, O, SC1
Nucleotides	
A	P, S, A1, A2, A3
U	P, S, U1, U2, U3
G	P, S, G1, G2, G3
C	P, S, C1, C2, C3

Table 1: Pseudo atoms for protein-RNA complexes. Each amino acid is represented by two main chain pseudo atoms (N and O) and maximum of two side chain pseudo atoms (SC1 and SC2). Each nucleotide is represented by five pseudo atoms, one each for phosphate and sugar molecules and three for the base. Gly has one extra main chain pseudo atom CA.

resented by the N and O atoms except in Gly where an additional main chain CA atom is used. In case of RNA chain only one pseudo atom used for phosphate and sugar molecules and three for each bases (Table 1).

A pairwise interaction is counted if the distance between pseudo atoms of protein and RNA is within 4.5 Å. We computed the frequency of all pairwise interactions for the whole 81 complexes and converted them into a contact potential using the following equation:

$$\mathbf{V(P_iN_j) = -RT \ln \frac{\Sigma_{81}(P_iN_j)}{\Sigma_{81}P_i * \Sigma_{81}N_j}}$$

Where P_iN_j is the observed frequency of a particular atom pair of protein and RNA that are within the cut-off distance given above, and P_i is the frequency of the i^{th} protein atom interacting with RNA atoms and N_j is the frequency of the j^{th} RNA atom interacting with protein atoms (in the data set). The contact potential for each pair represents the energy minimum or saddle point of a Lennard-Jones (LJ) type potential (as implemented in Attract, [7]). The minimum pairwise distance between pseudo atom pairs represents the effective contact radius in the LJ-potential.

3 Results and Discussion

The pairwise contact potential between two pseudo atoms of protein and RNA is shown in figure 1. Aromatic residues show no preference to interact with the sugar (S) or phosphate (P) groups in the nucleotide but the interaction with the nucleo-bases is very favorable with few exceptions. This is due to possible stacking interactions between the aromatic ring and bases that may help to stabilize a protein-RNA complex. However, no interaction between the side chain pseudo atoms of Trp and the pseudo atoms of Uracyl base were found. Similar to aromatic residues, aliphatic (hydrophobic) residues (Ile, Leu and Val) show a preferential interaction with the nucleotide bases. Positively charged residues Arg and Lys prefer to interact with the negatively charged phosphate groups but do not interact favorably

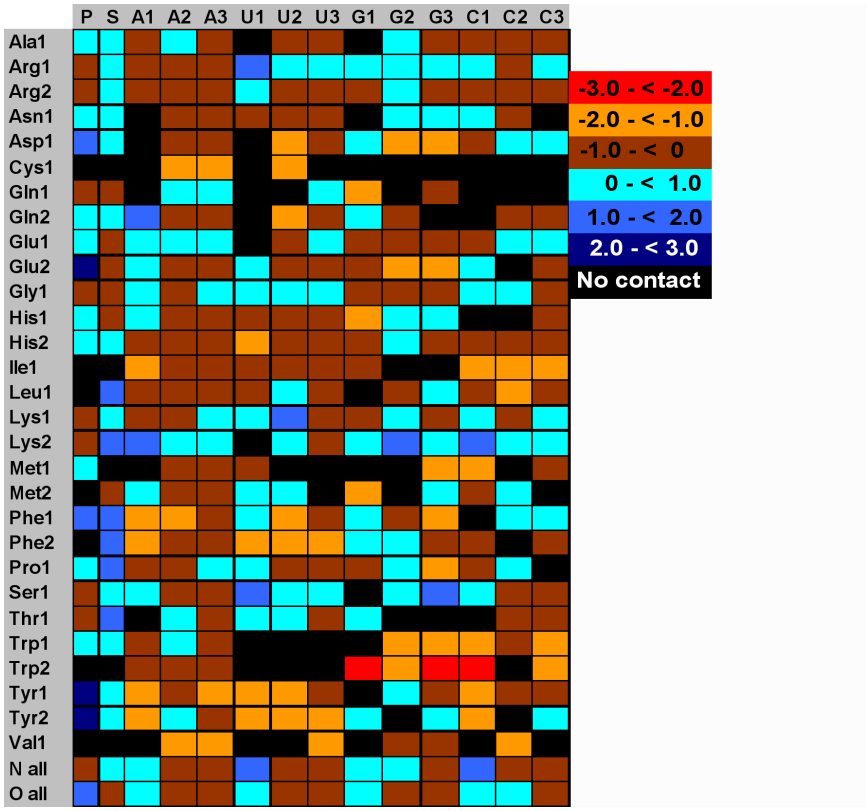


Figure 1. Pair-potential at the protein-RNA interfaces. Detail of the atom types is given in Table 1. If there is no interaction between an atom pair then that cell is colored black. Values should be multiplied by 10^{-3} in real scale.

with the ribose group. In addition, Arg interacts preferentially with the Adenine base but only moderately with the other three bases. Aspartic acid, being negatively charged, has less favorable interactions with P and S but it interacts favorably with the bases. Another negatively charged residue Glu has less preference to interact with P but more to S. Main chain atoms of amino acid residues have a mixed preference to interact with the nucleotide atoms. The knowledge-based potential has been integrated in the flexible docking program ATTRACT which employs energy minimization in translational and rotational degrees of freedom of the interacting partners [7]. Initial tests indicate that the potential can reproduce in many cases near-native protein-RNA complexes in good agreement with experimental complex structures (Figure 2).

4 Concluding Remarks

The approach was already used to predict a protein-RNA complex given in the CAPRI challenge [8] starting from two unbound structures which is under evaluation. We are

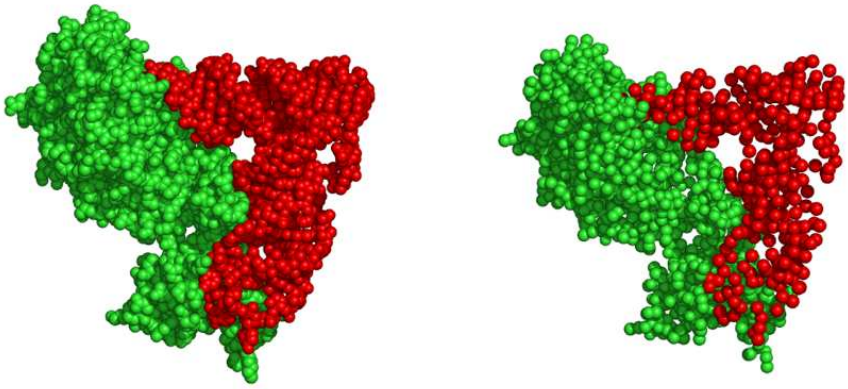


Figure 2. An all atom (left) and reduced model (right) of the Aspartyl tRNA synthetase complexed with tRNA (1asy).

now working on a set of known protein-RNA complexes to test the performance of the potential to predict the native complex structure starting from the individual components in systematic docking searches.

Acknowledgments

We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support through grant Za153-11 to M.Z.

References

1. K. Nadassy, S. J. Wodak and J. Janin J, *Structural features of proteinnucleic acid recognition sites*, Biochemistry **38**, 1999–2017, 1999.
2. S. Jones, D. Daley, N. Luscombe, H. Berman and J. Thornton J, *ProteinRNA interactions: a structural analysis*, Nuc. Acids Res. **29**, 943–954, 2001.
3. M. Treger and E. Westhof, *Statistical analysis of atomic contacts at RNA-protein interfaces*, J. Mol. Recog. **14**, 199–214, 2001.
4. J. J. Ellis, M. Broom and S. Jones, *ProteinRNA interactions: structural analysis and functional classes*, Proteins **66**, 903–911, 2007.
5. R. P. Bahadur, M. Zacharias M and J. Janin, *Dissecting protein-RNA recognition sites*, Nuc. Acids Res. **36**, 2705–2716, 2008.
6. H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, et. al, *The Protein Data Bank*, Acta Crystallog. Sect. D **58**, 899–907, 2002.
7. M. Zacharias, *Protein-protein docking with a reduced protein model accounting for side-chain flexibility*, Protein Sci. **12**, 1271–1282, 2003.
8. J. Janin, K. Henrick, J. Moult, L. T. Eyck, M. J. Sternberg, S. Vajda, I. Vakser, S. J. Wodak, *CAPRI: a Critical Assessment of PRedicted Interactions*, Proteins **52**, 2–9, 2003.

Fast and Accurate Structure-Based Calculation of Folding Free Energies and Binding Affinities

Caroline M. Becker[1], Alexander Benedix[1], Bert L. de Groot[2], Amedeo Caflisch[3], and Rainer A. Böckmann[1]

[1] Theoretical and Computational Membrane Biology, Center for Bioinformatics Saar, 66041 Saarbrücken, Germany
E-mail: {c.becker,benedix,rainer}@bioinformatik.uni-saarland.de

[2] Computational Biomolecular Dynamics Group, Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[3] Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

We present a fast structure-based method for the prediction of folding free energies and protein-protein binding affinities, including full flexibility. The method replaces molecular dynamics simulations by a fast generation of alternative protein-protein conformations based on geometric considerations only. The energy function is based on physical chemistry and an efficient continuum solvent approach. The correlation between experimental and predicted free energies obtained for a dataset including almost 600 mutants and more than 350 protein-protein complex mutants is ≈ 0.8 with a standard deviation of 1 kcal/mol[1]. Owing to its velocity and its predictive power, the method can be applied to complete mutational scans.

1 Introduction

Protein-protein interactions are involved in most processes in the cell and are therefore an important target in pharmaceutical research. By inhibition or increase of protein-protein complex formation the activity of many processes can be influenced. However, the directed design of protein interaction surfaces with defined properties involves large-scale mutational scans. Because of their inherent computational complexity, these are precluded by the most rigorous and accurate methods in this context, the free energy perturbation (FEP) and the thermodynamic integration (TI) methods. These make use of a physical effective energy function (force field). In these methods, also termed computational alchemy, integration over the free energy gradient with respect to a perturbation parameter yields the free energy difference between two states. Both methods require molecular dynamics simulations at least in the nanosecond time range to reach sufficient convergence, using explicit solvation. FEP and TI probably work best for conservative single point mutations, but have also been successfully applied to calculate absolute binding free energies. Hybrid methods like the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA[2]) method combine the calculation of molecular mechanics free energies with continuum solvent calculations.

Here, we developed a fast method, both to estimate the effect of mutations on the folding free energy of isolated proteins as well as on the protein-protein binding affinity. This fast structure-based prediction makes a systematic computational mutagenesis of protein interfaces feasible and thus will allow for a thorough analysis of the binding characteristics of

protein complexes, a pre-requisite for the design of mutants with specified properties, or for the determination of protein-ligand binding affinities. The proposed method combines a fast generation of conformations based on geometrical constraints only with a physical effective free energy function.

2 Methods

In our method[1], the conformational flexibility of proteins or protein-protein complexes is treated by CONCOORD[3]: Interactions in the (mutated) crystal input structure are analyzed and translated into geometrical constraints. Type-dependent margins are added on different interaction classes. Starting from random coordinates, the structure is iteratively rebuilt fulfilling all geometrical constraints. In this way, multiple independent structures are generated where each conformation is uncorrelated with previously generated ones[3]. For the prediction of free energies based on these (energy-minimized) structural ensembles, 300 structures were generated. The free energy is approximated by an energy function similar to the MM/PBSA approach[2]:

$$\Delta G = \Delta G_{\text{solvation}} + \Delta G_{\text{electrostatic}} + \Delta G_{\text{MM}} + \Delta G_{\text{entropy}} \; . \tag{1}$$

Figure 1 shows the used thermodynamic cycle to compute the effect of mutations on protein-protein binding affinity. The change in affinity is computed according to

$$\begin{aligned} \Delta\Delta G &= \Delta G^{\text{bind}}_{\text{Mutant}} - \Delta G^{\text{bind}}_{\text{Wildtype}} \\ &= \Delta G^{\text{mutate}}_{\text{complex}} - \Delta G^{\text{mutate}}_{\text{single}} \; . \end{aligned} \tag{2}$$

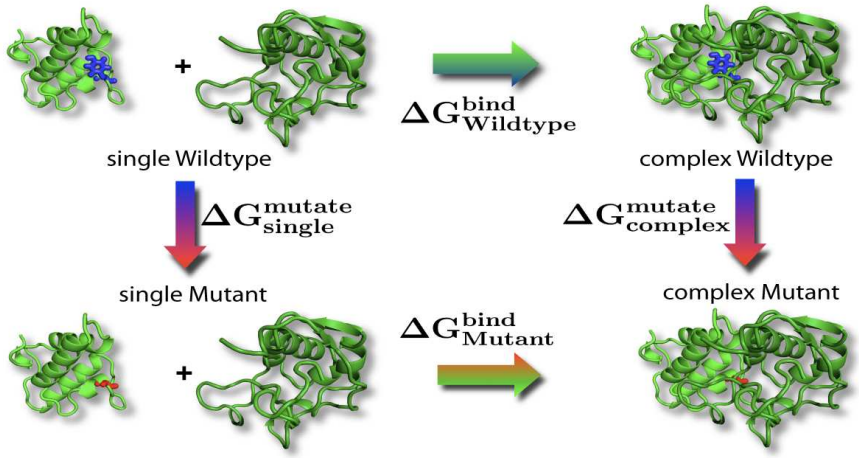


Figure 1. Thermodynamic cycle used to estimate the effect of a mutation on the protein-protein binding affinity.

3 Results

Figure 2 displays the flexibility of the response regulator protein of bacterial chemotaxis, CheY, upon the D12A mutation (highlighted in green). Especially the flexibility of neighboured residues is significantly enhanced for the mutant (lower right). The overall correlation achieved for the folding stability of almost 600 mutants is 0.75 with a standard deviation of less than 1 kcal/mol. For protein-protein binding, the obtained correlation is ≈ 0.8 with a standard deviation of 1.2 kcal/mol.

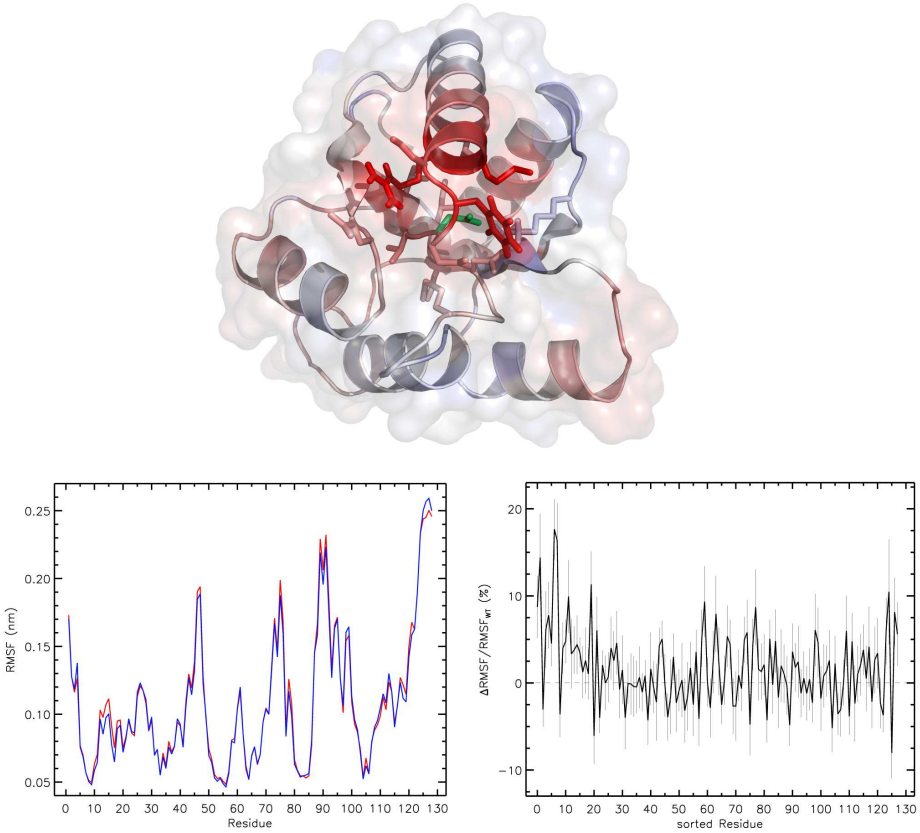


Figure 2. Color-coded change in flexibility of the response regulator protein of bacterial chemotaxis, CheY, upon the mutation D12A. The change in flexibility is additionally shown as a function of the residue number and as a function of the distance of the respective residues to the mutation site.

Acknowledgments

AB, CB, BLG, and RAB acknowledge support by the Deutsche Forschungsgemeinschaft (Grant Bo 2963/1-1, GR 2079/4-1, and BIZ 4/1). Part of the work was supported by the Swiss National Center for Competence in Research in Structural Biology (RAB and AC).

References

1. A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch, R. A. Böckmann, *CC/PBSA: Structure-based prediction of free energy and flexibility of mutants*, in preparation.
2. P. A. Kollman et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum methods*, Acc. Chem. Res. **33**, 889–897, 2000.
3. B. L. de Groot, D. M. F. van Aalten, R. M. Scheck, A. Amadei, G. Vriend, H. J. C. Berendsen, *Prediction of protein conformational freedom from distance constraints.*, Proteins **29**, 240–251, 1997.

Correlation Effects in Protein-Protein Recognition

Hans Behringer and Friederike Schmid

Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany
E-mail: {behringer, schmid}@physik.uni-bielefeld.de

Correlation effects in the distribution of hydrophobic and polar residues are investigated within an idealised coarse-grained model for the recognition of two rigid biomolecules such as proteins. To this end, a two-stage approach is adopted where the biomolecules are first optimised with respect to each other and afterwards their selectivity is tested in the presence of other molecules. Correlations lead to different optimum characteristic lengths of the hydrophobic and polar patches for the design of the two biomolecules on the one hand and their selectivity in the presence of other molecules on the other hand.

1 Introduction

Biomolecular recognition, that is the ability of a biomolecule to interact specifically with another molecule in an heterogeneous environment of structurally similar rival molecules, is an essential component in biological systems. The recognition process between two molecules is governed by a complicated interplay of non-covalent interactions of strengths comparable to the thermal energy[1]. This implies that the study of idealised models with methods from statistical physics might lead to valuable insight into this problem.

2 Model and General Approach

In this work we consider protein-protein recognition on a coarse-grained level in the framework of idealised models. The biomolecules are assumed to undergo no refolding during the association process which is a justified assumption for most protein-protein recognition processes[1]. Motivated by the observation that hydrophobicity is the major driving force in molecular recognition[1] we describe the type of the residue at the position $i = 1, \ldots, N$ of the interface by a binary variable $\sigma_i \in \{\pm 1\}$ for the target molecule and by $\theta_i \in \{\pm 1\}$ for the interaction partner[2]. We then model the energetics at the two-dimensional interface by

$$\mathcal{H}(\sigma, \theta; S) = -\varepsilon \sum_{i=1}^{N} \frac{1 + S_i}{2} \sigma_i \theta_i \tag{1}$$

as a direct contact interaction of strength ε. The variable S_i takes on the two values ± 1 and describes the local fit of the shape of the molecules at the interface resulting from a rearrangement of the amino acid side chains when the complex is formed[1].

To study the recognition process between two rigid biomolecules we adopt a two-stage approach. For a fixed target sequence $\sigma^{(\mathrm{t})} = (\sigma_1^{(\mathrm{t})}, \ldots, \sigma_N^{(\mathrm{t})})$ we first design an ensemble of probe molecules θ at a design temperature $1/\beta_{\mathrm{D}}$ leading to the distribution $P(\theta|\sigma^{(\mathrm{t})}) = \frac{1}{Z_{\mathrm{D}}} \sum_S \exp\left(-\beta_{\mathrm{D}} \mathcal{H}(\sigma^{(\mathrm{t})}, \theta; S)\right)$. In a second step the free energy difference of association at temperature $1/\beta$ is calculated for the interaction of the probe ensemble with the target molecule $\sigma^{(\mathrm{t})}$ and a structurally different rival molecule $\sigma^{(\mathrm{r})}$. In this step the free energy of

the interaction of the molecule $\sigma^{(\alpha)}$ with a particular probe sequence θ has to be averaged with respect to the distribution $P(\theta|\sigma^{(\mathrm{t})})$ giving finally the selectivity $\Delta F = F_{\mathrm{target}} - F_{\mathrm{rival}}$. A negative ΔF then signals recognition of the target.

For the majority of real protein-protein complexes the appearance of extended but fairly small patches of residues of the same type has been reported[3]. Biomolecular binding seems also to be strongly influenced by small-scale structures[4]. We therefore consider molecules which have correlated recognition sites at the interface with extended patches of residues of the same type. This can be incorporated into our model by adding additional correlation terms like

$$\mathcal{H}_{\mathrm{cor}} = -\gamma \sum_{\langle i,j \rangle} \theta_i \theta_j - \mu \sum_i \theta_i. \tag{2}$$

to the Hamiltonian of the system. The correlation parameters γ and μ (for the different types of molecules) are then used to fix the hydrophobicity and correlation length on the recognition sites. The average extension of the patches of residues from the same class is used as a measure for the correlation length of the finite system. Introducing the complementarity parameter $K = \sum_i \sigma_i \theta_i$ which measures the structural fit of the two biomolecules σ and θ at the interface, the selectivity averaged over all targets and rivals with the same correlation properties turns out to be the negative difference between the averaged complementarity with the target and the rival molecules, respectively[5].

3 Results

In this section molecular recognition for target and rival molecules with a fixed average hydrophobicity $h = 0.4$ per residue and a fixed (to length unity normalised) correlation length $\lambda = 0.263$ is considered within the model discussed above. The selectivity is studied as a function of the correlation length of the recognition site of the probe molecules. Figure 1

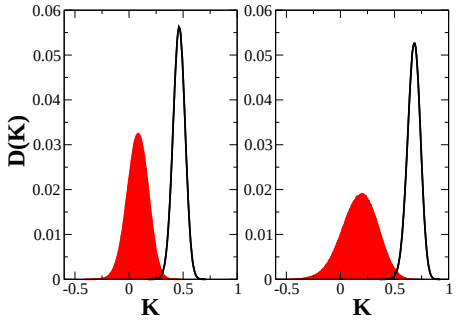


Figure 1. Distribution of the complementarity of the probe molecules with the target molecules (solid curve) and the rival molecules (shaded curve) for uncorrelated (left) and correlated probe molecules (with correlation length $\lambda_{\mathrm{p}} = 0.25$, right).

shows the distributions of the complementarities. For uncorrelated probe molecules the distribution for the complementarity with the target molecules is clearly separated from

the one with the rival molecules and shifted to larger values. This indicates the overall recognition ability of the system. A moderate increase of the correlation length on the probe molecules shifts the distribution to larger values of the complementarity so that an increased selectivity is expected. The first moments of the distributions are shown in figure 2. The average complementarity of the probe molecules with the target is always larger than that of the probe molecules with the rival. In the extreme cases where the correlation length tends to the minimum and maximum possible values the two averages become identical indicating that selectivity is lost as the probe molecules are not structured any more with respect to a particular molecule. The selectivity as shown in figure 2 has an optimum at a correlation length that is shifted below the value corresponding to the optimum of the complementarity with the target molecules. A smaller correlation length implies the appearance of an increased number of smaller patches on the recognition site of the probe molecule and hence an entropic profit for the interaction with the target due to more possible ways to align each other favourably. This effect only influences the contributions from the target-probe interactions due to the optimisation during the design step. The rival-probe interaction is not optimised and hence it is insensitive to a matching of structure elements.

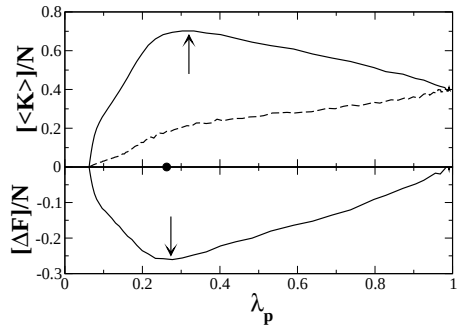


Figure 2. Upper part: Complementarity with the target molecules (solid curve) and the rival molecules (dashed curve) as a function of the (normalised) correlation length of the probe molecules (the fixed correlation length of the target and rival molecules is shown by the circle). Lower part: Resulting selectivity of the system.

Acknowledgments

We thank the Deutsche Forschungsgemeinschaft (SFB 613) for financial support.

References

1. C. Kleanthous, ed., *Protein-Protein Recognition* (Oxford University Press, Oxford, 2000).
2. H. Behringer, A. Degenhard, and F. Schmid, Phys. Rev. Lett. **97**, 128101, 2006.
3. T. A. Larsen, A. J. Olsen, and D. S. Goodsell, Structure **6**, 421, 1998.
4. T. Bogner, A. Degenhard, and F. Schmid, Phys. Rev. Lett. **93**, 268108, 2004.
5. H. Behringer and F. Schmid, in preparation.

Prediction of Twist of Amyloid Fibrils Using Molecular Dynamics

Josh Berryman[1], **Sheena Radford**[2], **and Sarah Harris**[1]

[1] School of Physics and Astronomy, University of Leeds, UK
E-mail: {j.berryman, s.a.harris}@leeds.ac.uk

[2] Institute of Molecular and Cell Biology, University of Leeds, UK
E-mail: s.e.radford@leeds.ac.uk

Many proteins and peptides form amyloid fibrils. These long, helically symmetric aggregates can be highly ordered but are not normally amenable to structure determination by X-ray crystallography or solution NMR. Therefore although amyloid fibrils of the same sequence can display substantial variation in gross morphological features such as twist (depending on seeding and on growth conditions such as pH) the atomic-level origins of this variation remain obscure. In order to probe the origins of the diversity in fibrillar twist we use the weighted-histogram analysis method of atomistic molecular dynamics (WHAM) to measure the free energy with respect to twist of two model polyalanine fibrils having different subunit-packing symmetries within the overall helical symmetry. We identify differences in twisting which arise from the differences in symmetry.

1 Introduction

The interest in amyloid formation arises from its involvement in a number of human diseases and from the potential of amyloid to serve as a nano-material if assembly could be sufficiently well controlled. A connection has been proposed between the free-energy of twisting of an amyloid sheet and the number of sheets which can be stacked together within a fibril[1], so a numerical method of calculating the free-energy of twisting is important for this reason as well as for the purpose of checking hypothetical atomic structures based on partial experimental data.

2 Materials and Methods

Two small model amyloid fibrils were chosen for this study, each composed of two β-sheets of six CH_3CO-$(Ala)_7$-$NHCH_3$ peptide strands. The model fibrils differed from each other in the symmetry of the arrangement of peptide strands: the first system (P) was composed of parallel β-sheets arranged face-to-face antiparallel to eachother (symmetry class 1 in the now-standard nomenclature[2]) and the second (AP) of antiparallel sheets arranged face-to-face (symmetry class 5).

Following previous studies of the free energy of DNA with respect to helical pitch[3], twist restraints were applied to adjacent peptide strands. Torsions were applied at 2nd, 4th and 6th C^α atoms of each strand in the pattern 2-4-4'-2' and 6'-4'-4-6, where prime (') indicates the C^α is on an axially adjacent strand within the sheet. WHAM[4] was applied, stepping in 2° increments from 0° to -16° and from 0° to $+16^\circ$, with a negative twist angle indicating left-handed helicity. Restraints had a strength of 750 $\mathrm{kcalmol}^{-1}\mathrm{rad}^{-2}$.

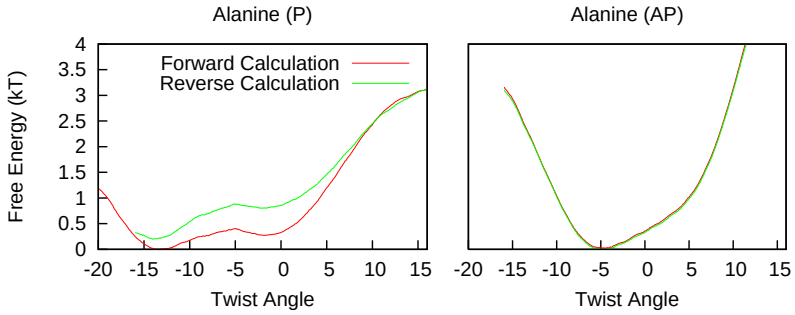


Figure 1. **Free energy with respect to twist.** Convergence is better for the AP system (reversibility is greater), and equilibrium twist is smaller. The P system shows a local minimum at a small twist, although the global minimum is strongly twisted. Energy is per torsion, shown are averages over the 12 central torsions.

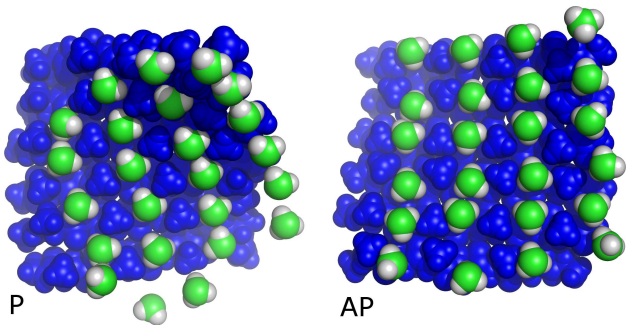


Figure 2. **Comparison of interfacial surfaces.** These snapshots of equilibrated unrestrained structures show the different characters of the two inter-sheet surfaces of the model fibrils. The back sheet is shown in blue, while only the inward-facing sidechains from the front sheet are shown (in green and white). The grooves in the AP surface are *compartmentalised*, preventing the sheets from sliding past eachother. This alters the twisting behaviour compared to the P system.

Measurements involving strands from either end of the fibril were discarded, in order to minimise contributions due to edge effects. Simulations were parameterised using the AMBER ff99-SB forcefield and run using the AMBER9[5] package, with a GBSA solvent model[6]. Unrestrained simulations were also run for 21ns, in order to verify the results. All simulations had additional distance restraints at the central hydrogen bond ladder to prevent them from dissociating, with a cuton of 5.5Å between the O and H atoms.

3 Results

Free energies with respect to twist are displayed for the P and AP systems in figure 1. By displaying results from both the forward (outward) and reverse (inward) simulations we can estimate the error in the calculation as the difference between them, which is modest for the P system and negligible for the AP system. Both systems twist; the P system twists more than the AP but also has a local minimum with less twist.

4 Concluding Remarks

Pioneering studies of β-sheet geometry[7] predicted a greater twist for antiparallel sheets. The tendency of our P systems to twist more than the AP systems therefore requires explanation. Assuming that both results are correct, the source of the difference must be the interface between the two sheets, as the original studies were of isolated sheets. Figure 2 shows that the sheet-sheet interface for AP amyloid is more complex than for P, having a knobs-and-holes character. This leads to a diminished flexibility away from the rectangular conformation, thus limiting twist.

The presence of a local minimum for the twist of the P system suggests that variation in twist of experimentally observed fibrils could arise not only from variation in packing symmetry but also potentially from kinetic trapping in states which have a twist other than that of the global minimum.

Acknowledgments

We are endebted to M. Zacharias, S. Kannan and K. Kolhoff for inspiration and assistance. Alan Grossfield's WHAM code was used. This work was funded by the EPSRC.

References

1. A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov, and N. Boden, *Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta -sheet tapes, ribbons, fibrils, and fibers*, Proc. Natl. Acad. Sci. USA, **98**, no. 21, 11857–11862, 2001.
2. M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel, and D. Eisenberg, *Atomic structures of amyloid cross-β spines reveal varied steric zippers*, Nature, **447**, no. 7143, 453–457, 2007.
3. S. Kannan, K. Kohlhoff, and M. Zacharias, *B-DNA Under Stress: Over- and Untwisting of DNA during Molecular Dynamics Simulations*, Biophys. J., **91**, no. 8, 2956–2965, 2006.
4. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*, J. Comput. Chem., **13**, no. 8, 1011–1021, 1992.
5. D.A. Case et al., *AMBER9*, University of California, San Francisco, 2006.
6. G. D. Hawkins, C. J. Cramer, and D. G Truhlar, *Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium*, J. Phys. Chem., **100**, no. 51, 19824–19839, 1996.
7. K.-C. Chou, M. Pottle, G. Nemethy, Y. Ueda, and H.A. Scheraga, *Structure of β-sheets : Origin of the right-handed twist and of the increased stability of antiparallel over parallel sheets*, J. Mol. Biol., **162**, no. 25, 89–112, November 1982.

Automatic Sequential NOESY Assignment and NMR Structure Improvement by X-Ray

K. Brunner, W. Gronwald, A. Fischer, J. Trenner, K.-P. Neidig, and H. R. Kalbitzer

Institute of Biophysics and Physical Biochemistry, University of Regensburg,
93040 Regensburg, Germany
E-mail: konrad.brunner@biologie.uni-regensburg.de

and Bruker Biospin, 76287 Rheinstetten, Germany

We are developing AUREMOL[1] (`www.auremol.de`), which goal is the reliable and automatic structure determination of biological macro molecules such as proteins from NMR data. For a fully automatic sequential NOESY assignment the tool ASSIGN[2] has been developed. The required input consists of a homologous structure for a NOESY spectrum simulation and the experimental NOESY spectrum. ASSIGN fits the simulated NOE signals to the experimental spectrum. The fit quality given by a probability depends on the line shapes and volumes of the signals. The assignment is varied by moving or swapping spin system assignments using a Monte Carlo approach. A threshold accepting algorithm (TA[3]) is employed to find the maximum of accordance.

1 Introduction

For a fully automated sequential NOESY assignment the tool ASSIGN has been developed. The assignment is driven by the comparison of experimental spectra of a protein and simulated spectra. The simulated spectra are derived from a preliminary structure model. ASSIGN is part of the AUREMOL NMR software suite.

2 Method

The basic idea is to use preliminary structural information together with the NOESY peaks to drive the assignment process. Therefore ASSIGN expects additionally to the NOESY spectra a preliminary structure model of the protein to be solved. Such a model can be provided for example by homology modelling[4]. A start assignment can be provided as an optional input. The first step is the recording and the processing of a NOESY spectrum. In this spectrum the signals are identified and the corresponding chemical shifts are stored in a slot list. The second step is to simulate a NOESY spectrum for the structure. Each expected coupling signal is simulated with a proper line shape and volume. The shifts for the signals are not calculated. Instead of that shifts are taken from the slot list and randomly assigned to the simulated signals. If a start assignment is provided as input the shifts are assigned according to the start assignment. In the third step the resulting simulated spectrum is compared with the experimental one with respect to line shapes and signal volumes. The degree of accordance is expressed as a probability. In the following a quenching protocol is applied to the simulation procedure to improve the agreement of the spectra. A random perturbation swaps the shifts of two simulated signals and the probability of accordance is recalculated. If the new parameters lead to an improved agreement with the experimental

data, they are accepted, otherwise declined. This method is repeated until the agreement between experiment and simulation cannot be further improved for a defined number of iterations. As a result a sequential shift chemical assignment is obtained that can explain the experimental spectra with the final probability of accordance.

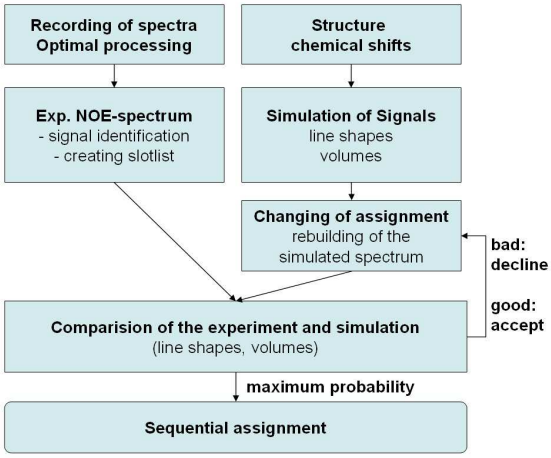


Figure 1. Scheme of the ASSIGN algorithm.

The agreement with the experimental data is expressed as probability and calculated on each test region where an experimental signal is found. Shapes are compared with the help of the cosinus criterion, volumes are compared directly by summing up the intensities of the testing regions. With the help of frequency distributions of solutions with random and partial correct assignments probabilities of line shape and volumes are derived. For the test region p the Bayesian probability of shape and volume (PSV) is given as

$$PSV_p = \frac{P_p^{S,ok} P_p^{V,ok}}{P_p^{S,ok} P_p^{V,ok} + P_p^{S,rnd} P_p^{V,rnd}} \tag{1}$$

$P^{S,ok}$ and $P^{V,ok}$ are the probabilities of the partial correct solution and $P^{S,rnd}$ and $P^{V,rnd}$ are the probabilities of the random solution. The sum over all (N_{ex}) test regions ESV (Energy Shape Volume)

$$ESV = \sum_{p=1}^{N_{ex}} |\ln(PSV_p)| \tag{2}$$

is optimized by the TA-algorithm[3].

3 Results

In order to evaluate the method, a set of test cases is considered based on the structure of HPr *S. aureus* (H15A) that has already been solved by NMR. In case of an ideal data

set the simulated spectrum is calculated from the known structure and is also used as an experimental spectrum. Here 450 of 455 shifts (99.3 %) were found correct without any partial start assignment. In case of a real experimental spectrum of HPr *S. aureus* (H15A) from up to 20 % correct start assignments over 90 % to 99 % correct assignments were found (Fig. 2).

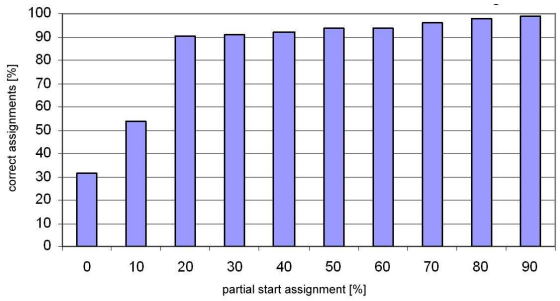


Figure 2. Correct assignments against partial start assignment.

In a third test case only easily obtainable NMR data such as chemical shifts of the backbone atoms H^N and H^α is used. In HPr *S. aureus* (H15A) these are 36.5 % of the whole assignment. With this start assignment 500 structures were calculated. The 10 best in respect of energy were taken for further analysis (Fig. 3B). Then ASSIGN was used to assign the missing shifts of the side chain atoms. 85.2 % of the shifts were correctly found. The structure bundle calculated from these data is shown in Fig. 3C. A clear improvement of the structure can be seen easily and the bundle is very similar to the original bundle (Fig. 3A). Also the AUREMOL NMR R-factor[5] (0.589 → 0.328), the RMSD (0.151 nm → 0.015 nm) and the Ramachandran[6] values are improved.

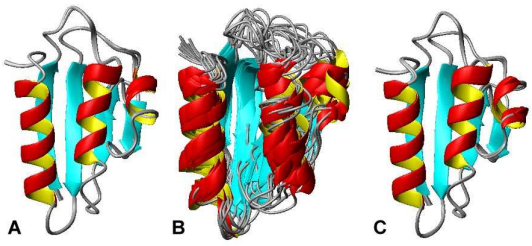


Figure 3. (A) Original structure, (B) structure calculated from bb-atoms, (C) ASSIGN improved structure of HPr *S. aureus* (H15A).

In the last test case the sequential assignment of the mutant HPr *S. aureus* (H15A) should be found with help of the solved structure of HPr *S. aureus* (wt). In this example 79.8 % of the assignments were already given (Fig. 4B). ASSIGN finds 96.7 % of the

correct assignments, which leads to the structure seen in Fig 4C. Again quality values such as AUREMOL R-factor (0.384 → 0.356), RMSD (0.046 nm → 0.027 nm) and the Ramachandran are improved.

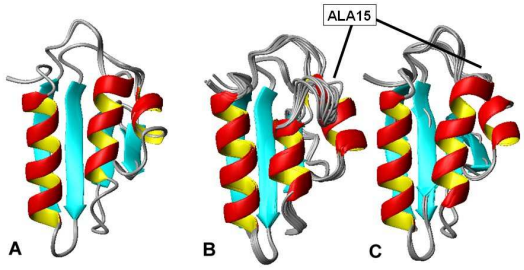


Figure 4. (A) Original structure, (B) structure calculated from the sequential assignment from the wt, (C) ASSIGN improved structure of HPr *S. aureus* (H15A).

For NMR structure improvement by X-ray data please see the ISIC algorithm[7].

Acknowledgments

Financial support from the BMBF, the Deutsche Forschungsgemeinschaft (DFG), and the European Union is gratefully acknowledged.

References

1. Gronwald, W. and Kalbitzer, H. R., *Automated Structure Determination of Proteins by NMR Spectroscopy*, Progr. NMR Spectr. **44**, 33–96, 2004.
2. Brunner, K., *Modellierung, Strukturverbesserung und sequentielle Zuordnung als vollautomatische Module für die automatisierte Proteinstrukturbestimmung im Softwareprojekt AUREMOL*, (Dissertation, Universität Regensburg, Germany, 2006).
3. Dueck,G. & Scheuer,T., *Threshold accepting: A general purpose algorithm appearing superior to simulated annealing*, J. Comput. Phys. **90**, 161–175, 1990.
4. Möglich,A., Weinfurtner,D., Gronwald,W., Maurer,T., & Kalbitzer,H.R., *PERMOL: Restraint-Based Protein Homology Modeling Using DYANA or CNS*, Bioinformatics **21**, 2110–2111, 2005.
5. Gronwald W., Kirchhofer R., Gorler A., Kremer W., Ganslmeier B., Neidig K.P., Kalbitzer H.R., *RFAC, a program for automated NMR R-factor estimation*, J Biomol NMR. **Jun;17(2)**, 137-51, 2000.
6. Laskowski,R.A., MacArthur,M.W., & Thornton,J.M., *Validation of protein models derived from experiment*, Curr. Opin. Struct. Biol. **8**, 631–639, 1998.
7. Brunner K., Gronwald W., Trenner J.M., Neidig K.P., Kalbitzer H.R., *A General Method for the Unbiased Improvement of Solution NMR Structures by the Use of Related X-Ray Data, the AUREMOL-ISIC Algorithm.*, BMC Struct Biol. **Jun 26;6(1):14**, 2006.

Exploring the First Steps of Aβ16-22 Protofibril Disassembly by N-Methylated Inhibitors

Yassmine Chebaro and Philippe Derreumaux

Laboratoire de Biochimie Théorique, UPR 9080 CNRS,
Institut de Biologie Physico-Chimique, Université Paris 7 Denis Diderot,
13 rue Pierre et Marie Curie, 75005 Paris, France
E-mail: {yassmine.chebaro, Philippe.Derreumaux}@ibpc.fr

Alzheimer's disease is characterized by the self-assembly of the Aβ(1-40)/(1-42) peptides. The design of efficient inhibitors is particularly challenging because the structures of the toxic Aβ species are transient in character and the modes of action of current inhibitors on Aβ oligomers are unknown. We know that N-methylated Aβ16-22 peptides (mAβ16-22) effectively inhibit fibrillogenesis and disassemble existing fibrils in vitro. In this work we report molecular dynamics (MD) and replica exchange molecular dynamics (REMD) simulations using the coarse-grained OPEP force field on a preformed protofibril of six Aβ16-22 peptides with either four copies of Aβ16-22 or four copies of mAβ16-22. While MD trajectories of 100 ns do not reveal any significant differences between the two systems, REMD simulations help understand the first steps of Aβ16-22 protofibril disassembly by N-methylated inhibitors.

1 Introduction

Alzheimer's disease is a fatal neurodegenerative disease characterized by the aggregation of amyloid-β peptides. There is increasing evidence that the oligomers of Aβ are themselves cytotoxic[1] and the hydrophobic region 16-22 is critical for aggregation. Several inhibitors have been designed in order to inhibit the aggregation process[2]. N-methylated inhibitors such as mAβ16-22 are known to prevent the growth of the fibrils and disassemble fibrils. However the mechanism of action of this inhibitor is not determined. Our work consists on studying the early steps of the inhibition using a preformed protofibril in interaction with either Aβ16-22 and mAβ16-22 (called in the text Aβ and mAβ) by MD and REMD simulations. Our results show that N-methylated peptides disassemble the preformed fibril by three mechanisms.

2 Materials and Methods

We use the coarse grained force field `OPEP` (*Optimized Potential for Efficient peptide-structure Prediction*) in its last version[3], which describes the short-range as well as long-range interactions of proteins, in a reduced representation. The initial structure of all our simulations is shown in figure 1 : it consists of a preformed bi-layer of 6 Aβ, and 4 free Aβ or mAβ. In mAβ, the (-N**H**) group of residues (`Leu 17`, `Phe 19` and `Ala 21`) is replaced by (-N**CH3**). Each system is minimized and equilibrated at the desired temperature for 1 ns. MD simulations[4] of 100 ns are realized at constant temperature (330 K), using Berendsen's thermostat and an integration time step of 1fs. For the REMD simulations, we used 16 replicas for temperatures ranging from 280 to 440 K, exchanges between neighboring replicas are attempted every 7.5 ps. Periodic boundary conditions are using for the

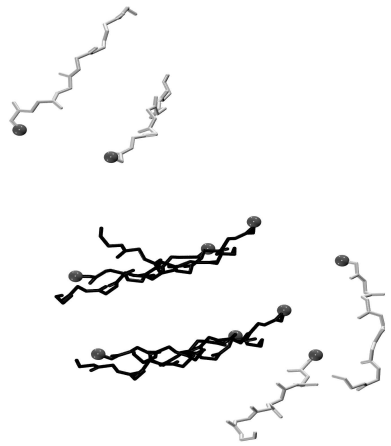


Figure 1. The starting structure for simulations : the preformed Aβ bi-layer (in black) in interaction with four copies of Aβ or mAβ (in gray). The N-terminal end of each peptide is indicated with a sphere.

systems, leading to a concentration of around 50 μM. The time of the REMD simulations is of 30 ns, noting that our purpose is to characterize the first steps of the inhibition mechanism, and not the thermodynamical behavior of the systems under equilibrium, requiring much longer simulation times. For the analysis, we calculate the nematic order parameter[5] P_2on the 6 peptides of the preformed fibril F1-F6, if $P_2 = 1$ then the system is fully parallel or antiparallel, otherwise if $P_2 = 0$ it is disordered.

3 Destabilization of the Fibril by N-Methylated Peptides

100 ns MD simulation of the system 10Aβ16-22 shows an average Cα-RMSD (calculated on the starting structure) of 2.7 Å for the first layer and 1.8 Å for the second one. Identical values are obtained for the simulation of 6Aβ + 4mAβ. Similarly no differences are found for the P2 values and the β-sheet content between the two systems, providing strong evidence that within the timescale explored, no difference in the aggregation process can be identified upon N-methylation. This is consistent with simulations using mAβ16-20[6]. To overcome the sampling encountered by MD simulations, we studied the effect of N-methylation by REMD. Figure 2.a shows the number of clusters and the P2-value averaged at every temperature for both the systems studied. These results show an increase of 32 % of the number of clusters and a decrease of 36.7 % of the P2-value in the presence of mAβ.

4 Binding of the N-Methylated Peptides to the Protofibril

In order to characterize at atomic level the effect of the inhibition upon the preformed fibril, we realized a clusterisation at all tempertaures using a RMSD cutoff of 3 Å on the six peptides F1-F6. These results show that the N-methylated peptides desorganize the protofibril at every temperature. Looking at the structures obtained at 280 and 315 K, we identified three mechanisms of binding between the N-methylated peptides and Aβ. Panels

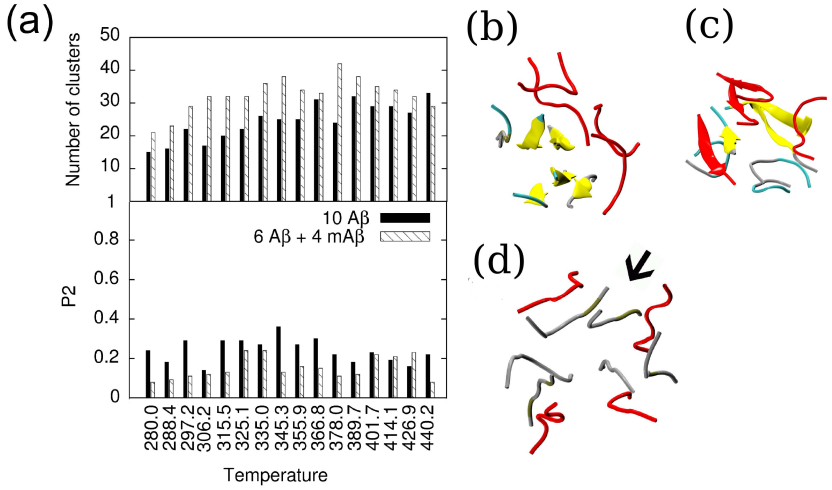


Figure 2. (a) Temperature-dependant properties of the peptides F1-F6 in 10 Aβ and 6Aβ+4mAβ simulations. (b,c) Center of the most populated clusters at 280 K for 10Aβ (b) and 6Aβ+4mAβ (c). (d) Center of one cluster at 315 K for 6Aβ+4mAβ. The N-methylated peptides are shown in red, the six preformed peptides F1-F6 use the following color code : yellow for β-sheet, grey for coil and blue for turn.

2.b and 2.c show the centers of the most populated clusters C1 with and without mAβ at 280 K. The Cα-RMSD calculated with respect to the preformed fibril are of 2.7 and 8.5 Å, respectively. In the structure of C1 in 2.c, two modes of binding are observed : on the top of an Aβ-sheet and a β-sheet formed between mAβ and Aβ peptides. Panel 2.d shows one cluster of the 6Aβ+4mAβ simulation at 315 K, presenting the third mode of interaction characterized by one intercalated mAβ and two Aβ peptides sequestered by two mAβ (see arrow in figure 2.d).

References

1. D.M. Walsh, et al. *Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo*, Nature **416**, 535-539, 2002.
2. D. J. Gordon, et al. *Inhibition of beta-amyloid(40) fibrillogenesis and disassembly of beta-amyloid(40) fibrils by short beta-amyloid congeners containing N-methyl amino acids at alternate residues*, Biochemistry **40**, 8237-8245, 2001.
3. J. Maupetit, P.Tuffery & P. Derreumaux. *A coarse-grained protein force field for folding and structure prediction*, Proteins **69**, 394-408, 2007.
4. P. Derreumaux & N. Mousseau. *Coarse-grained protein molecular dynamics simulations*, J Chem Phys **126**, 2, 2007.
5. P.H. Nguyen, M.S. Li, G. Stock, J.E. Straub & D. Thirumalai . *Monomer adds to preformed structured oligomers of abeta-peptides by a two-stage dock-lock mechanism*, Proc Natl Acad Sci USA **104**, 111–116, 2007.
6. P. Soto, M.A. Griffin & J.E. Shea. *New insights into the mechanism of Alzheimer amyloid-beta fibrillogenesis inhibition by N-methylated peptides*, Biophys. J. **93**, 3015-3025, 2007.

Simulation of the Outer Membrane Protein X in a Lipid Bilayer and in a Micelle

Alexandra Choutko[1], **Alice Glättli**[1,2], **and Wilfred van Gunsteren**[1]

[1] Eidgenössische Technische Hochschule Zürich, Physical Chemistry, 8093 Zürich, Schweiz
E-mail: {choutkoa, wfvgn}@igc.phys.chem.ethz.ch

[2] BASF SE, 67056 Ludwigshafen, Germany

The outer membrane protein X (OmpX) from *Escherichia coli* was simulated embedded in a phospholipid bilayer and as a protein-micelle aggregate. The resulting simulation trajectories were analyzed in terms of structural and dynamical properties of the membrane protein (MP). In agreement with experimental observations it was found that the β-barrel region, embedded in the lipophilic phase, is very stable, whereas the extracellular, protruding β-sheet, that plays an important role in cell adhesion and invasion of gram negative bacteria, shows large structural fluctuations. Additionally, water permeation into the core of the β-barrel protein was investigated: a very stable salt-bridge and hydrogen-bond network exists in that barrel and a water flux is therefore unlikely. No great difference in protein stability and dynamics between the bilayer and the micellar systems were observed.

1 Introduction

Membrane proteins (MPs) fulfill a wide spectrum of biological functions and it is estimated that MPs constitute 30 % of all proteins in living organisms. Furthermore, numerous diseases are directly related to MPs and more than 70 % of all currently available drugs are estimated to act via MPs. However, compared to water soluble proteins, structural information for MPs is sparse. Molecular dynamics (MD) simulation offers the possibility to describe the dynamic behavior of MPs. Additionaly, it is possible to compare the dynamics of a MP in the experimental environment, i.e. in protein-detergent co-crystals in X-ray diffraction experiments and in protein-micele aggregates in NMR solution-experiments, to its dynamics when embedded in a lipid bilayer.

OmpX belongs to the family of outer membrane proteins (Omp) of gram-negative bacteria and represents one of its smaller members. Its structure was studied by X-ray crystallography[1] and in a mixed protein-lipid micelle by TROSY NMR spectroscopy[2]. OmpX is characterized by eight antiparallel β-strands connected by three periplasmic turns and four extracellular loops (Figure 1). Four of the eight β-sheets protrude into the extracellular space, thus this β-sheet, also denoted as "waving flag", was suggested to act as a hydrogen bonding partner to other proteins in the extracellular space with complementary β strands at their surfaces[1]. This structural property of OmpX confirms conclusions drawn from microbiological studies that OmpX plays a key role in cell adhesion and mammalian cell invasion.

Here, OmpX was studied by MD when embedded in a lipid bilayer and as a protein-micelle aggregate. Experimental observations concerning the flexibility of the protruding β-sheet, the degree of stability of the hydrogen-bond network in the interior of the protein and the (in-)ability of OmpX to act as water pore can be ideally complemented with simulation studies at the atomic level.

2 Simulation Setup

The OmpX protein was simulated in a lipid bilayer, consisting of dimyristoyl-phosphatidylcholine (DMPC) molecules, and in a dihexanoyl-phosphatidylcholine (DHPC) micelle. The NMR structure[2] was employed as a starting structure. Two simulations, where OmpX was inserted in a lipid bilayer (simulations OmpX-DMPC-1 and OmpX-DMPC-2) and one simulation of the micellar protein-lipid aggregate (simulation OmpX-DHPC) were performed. The protein in the simulation OmpX-DMPC-1 initially contains no water in the internal cavities of OmpX, while in OmpX-DMPC-2, as well as in OmpX-DHPC, these cavities were initially filled with water (Table 1). All simulations were performed using the GROMOS simulation software[3] and the GROMOS biomolecular force field (version 45A3_CH95)[3,4].

system/simulation name	OmpX-DMPC-1	OmpX-DMPC-2	OmpX-DHPC
lipid and assembly type	DMPC, bilayer	DMPC, bilayer	DHPC, micelle
number of lipids	104	104	82
number of water molecules	6518	6559	12682
number of counter-ions	2 Na^+	2 Na^+	2 Na^+
total number of atoms	25876	25999	42044
type of box	rectangular	rectangular	truncated octahedron
box size [nm^3]	$5.8\times6.3\times9.2$	$5.8\times6.3\times9.5$	9.9–9.9–9.9
equilibration time [ns]	0.44	0.44	0.3
simulation (production) time [ns]	15	25	25

Table 1. Details of the simulation setups.

3 General Structural Analysis

In all three simulations, the root-mean-square-deviation (RMSD) of the protein from the initial NMR structure is rather high. This high RSMD is mainly affected by the extracellular loops, which are completely exposed to the solvent and show very large root-mean-square-fluctuations (RMSF). In contrast, the β-barrel and the periplasmic turns appear to be rather stable.

61 to 64 % of OmpX residues are on average observed to adopt a β-strand-like conformation in the simulations, while in the X-ray structure the β-strand content is higher (78%). The difference mainly originates from the the larger RSMF of the protruding β-sheet of OmpX in the simulations; while in the crystal, this β-sheet is involved in interactions with a neighboring OmpX molecule. In contrast, the barrel behaves as a rather rigid entity in the simulations as well as in the X-ray and NMR experiments.

4 Sturdy β-Barrel and Flexible "Waving Flag"

The overall agreement between NOE distances inferred from simulation and experiment is satisfactory. Simulations fulfill 78-85 % of the experimentally derived NOE distances and 52-60 % of all violations do not exceed 0.1 nm. In the extracellularly protruding β-sheet regions, only a few interstrand NOEs could be unambiguously assigned, and fast amide proton exchange was observed in the NMR solution experiments[2]. The "waving flag" seems therefore to be experiencing a large plasticity and a local fraying. The relatively large isotropic atomic B-factors derived from the X-ray diffraction data[1] of the protruding β-sheet are consistent with the NMR derived observations. However, the few NOE distances available for this region are not very well reproduced in the simulations, indicating that the protruding β-sheet region is somewhat too mobile in the simulations. The region of this four β-strands involved in the protruding β-sheet might show frequent transitions between hydrogen-bonded folded structure and solvent-exposed less secondary defined structure.

5 Water Exchange in the β-Barrel

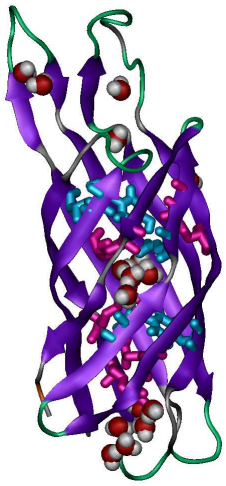


Figure 1. OmpX protein after a simulation of 25 ns in a micelle. The different colors indicate the secondary structure assignment: β-sheets are in violet, turns are in green, random coils are in white. Residues making hydrogen bonds to another residue for more than 45 % and more than 85 % of the simulation time are drawn respectively in blue and in pink. Water molecules are represented as red van der Waals spheres with grey hydrogens.

Most side-chains of polar and charged residues of OmpX point to the interior of the β-barrel and form a network of rather stable hydrogen bonds and salt-bridges (Figure 1). Consequently, no pathway exists between the extracellular and the periplasmic end of the barrel, making it unlikely to observe a continuous water flux through the barrel. Nevertheless, in the simulations of OmpX-DMPC-2 and OmpX-DHPC, some exchange between the internal water cavities and bulk water is observed.

6 Bilayer Versus Micelle Environment

From the data presented here, no significant differences in protein stability or dynamics can be detected between the simulations in a DMPC bilayer and a DHPC micelle. However, when comparing the few interstrand NOEs of the protruding β -sheet, the protein-micelle simulation appears to fulfill the long-range NOEs better than the simulations of the protein-bilayer system. The extracellular loops seem as well to have even more structural freedom in a bilayer system than in a micellar system.

Acknowledgments

We thank Dr. C. Fernandez for providing us with the NMR data. Financial support was obtained from the National Centre of Competence in Research (NCCR) Structural Biology of the Swiss National Science Foundation, which is gratefully ackowledged.

References

1. J. Vogt and G.E. Schulz, *The Structure of the Outer Membrane Protein OmpX from E. coli Reveals Possible Mechansims of Virulence*, Structure. **7**, 1301–1309, 1999.
2. C. Fernandez, C. Hilty, G. Wider, P. Güntert, and K. Wüthrich, *NMR Structure of the Integral Membrane Protein OmpX*, J. Mol. Biol. **336**, 1211–1221, 2004.
3. W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, (vdf Hochschulverlag, ETH Zürich, Switzerland, 1996).
4. I. Chandrasekhar, M. Kastenholz, R.D. Lins, C. Oostenbrink, L.D. Schuler, D.P. Tieleman, and W.F. van Gunsteren, *A Consistent Potential Energy Parameter Set for Lipids: Dipalmitoylphosphatidylcholine as a Benchmark of the GROMOS 45A3 Force Field*, Eur. Biophys. J. **32**, 67–77, 2003.

OPERA: An OPtimized coarse-grained Energy model for RnA

Claire Colas[1], **Phuong Hoang Nguyen**[2], **Jean-Christophe Gelly**[1], **and Philippe Derreumaux**[1]

[1] Laboratoire de Biochimie Théorique, UPR 9080 CNRS,
Institut de Biologie Physico-Chimique, Université Paris 7 Denis Diderot,
13 rue Pierre et Marie Curie, 75005 Paris, France
E-mail: Philippe.Derreumaux@ibpc.fr

[2] Institute of Physical and Theoretical Chemistry, J.W. Goethe University,
Max-von-Laue-Strasse 7, 60438 Frankfurt, Germany

RNAs have many cellular functions ranging from transcription to catalysis. The gap between sequences and 3D structures is increasing and knowledge of RNA dynamics and thermodynamics at an atomic level is missing. In principle, all-atom molecular dynamics (MD) and replica exchange molecular dynamics (REMD) simulations in explicit solvent can investigate these issues. However with current computer facilities, these simulations have been limited to small RNAs. To move to large RNAs, we can resort to coarse-grained models. In this study we present OPERA, a generic coarse-grained model for RNA. We report MD and REMD simulations on two RNAs of 22 nucleotides using a set of non-optimized OPERA parameters. Current results suggest that further optimization of the OPERA force field should open the door to a relevant model for studying large RNA such as riboswitches.

1 Introduction

All-atom molecular dynamics simulations in explicit solvent are often used to investigate the dynamics and thermodynamics of biomolecules, but they are time-consuming and slow to converge to equilibrium. Pande's group for instance managed to fold a 12 nucleotide RNA with Folding@home, but used 150 000 CPUs[1].

Reducing the number of degrees of freedom is one solution to accelerate convergence. Coarse-grained models have long history for proteins, but only 3 models exist for nucleic acids : one for DNA[2] and two for RNA[3,4], with each nucleotide represented by 3 beads : phosphate, sugar, base. Two of them are based on the G$\bar{o}$ potential, which requires knowledge of the native structure. The last model developed by Dokholyan *et al.* shows promising results combining DMD and a square-well potential, but is not free of any biases[4]. We present here our *ab initio* OPERA force field.

2 OPERA

In OPERA, each nucleotide is represented by 6 to 7 beads (Fig. 1): 1 bead for the phosphate, 4 beads for the sugar - O5', C5', C4', C1'- and 1 bead for the base in pyrimidines and 2 in purines. The number of particles is reduced of 80%. Three of the six torsional angles are thus conserved : α, β and γ. Note that the OH group specific to RNA is not treated explicitly and the sugar pucker cannot be modeled. The solvent is treated implicitly.

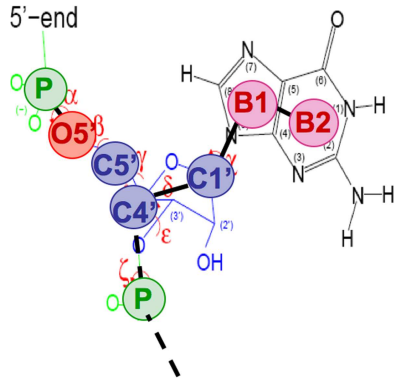


Figure 1. Our coarse-grained model is represented for a guanine superposed on an all-atom representation.

OPERA is based on the OPEP force field (Optimized Potential for Efficient peptide-structure Prediction) developed in our group for proteins[5,6]. The potential describes short range and long range interactions as follows : $E = E_{bonds} + E_{angles} + E_{torsion} + E_{LJ} + E_{HB}$. E_{LJ} is a 6-12 potential and E_{HB} is represented by two-body and four body interaction terms[7]. The geometrical parameters were derived from a statistical study of 220 structures of the PDB. Here, we use a set of non-optimized force constants.

3 Preliminary Results and Conclusions

MD simulations of 50 ns were performed at 310 K on two RNAs, each of 22 nucleotides : one hairpin (1EOR) and one pseudoknot (2G1W). Figures 2a and 2b show the RMSD of both systems with respect to their NMR structures. The hairpin (Fig. 2a) is rather stable, with the RMSD remaining around 4.5 Å during the first 32 ns and then around 3.5 Å. On the other hand, Figure 2b shows that the pseudoknot deviates much more. The RMSD evolves between 4 and 7 Å until 36 ns when it suddenly decreases to 2.5 Å (Fig. 2c) before reaching 11 Å at 50 ns.

REMD simulations were also performed on the hairpin, using 14 replicas with T ranging from 310 K to 360 K. Each replica is simulated for 150 ns and exchange events are attempted every 10 ps. Cluster analysis was done at 310 K using the 20-150 ns time interval with a RMSD cutoff of 2.5 Å. A total of 33 clusters is found and two clusters represent 75% of all conformations. The centers of these clusters are at 14.9 Å and 4.6 Å from the NMR structure. Figure 2d superposes the NMR structure on the center of the second cluster. Note that this state does display three non native H-bonds.

We have presented a new coarse-grained model for RNA. Our preliminary results, using a set of non-optimized force constants, are very encouraging since the hairpin is apparently stable at 310 K and the pseudoknot is in equilibrium between native-like and unfolded states. Analysis, based on 150 ns REMD simulations, confirms that the hairpin can adopt a native-like structure, albeit with a lower probability than an unfolded state. We believe that optimization of OPERA should stabilize the structure and capture the thermodynamics of large RNAs.

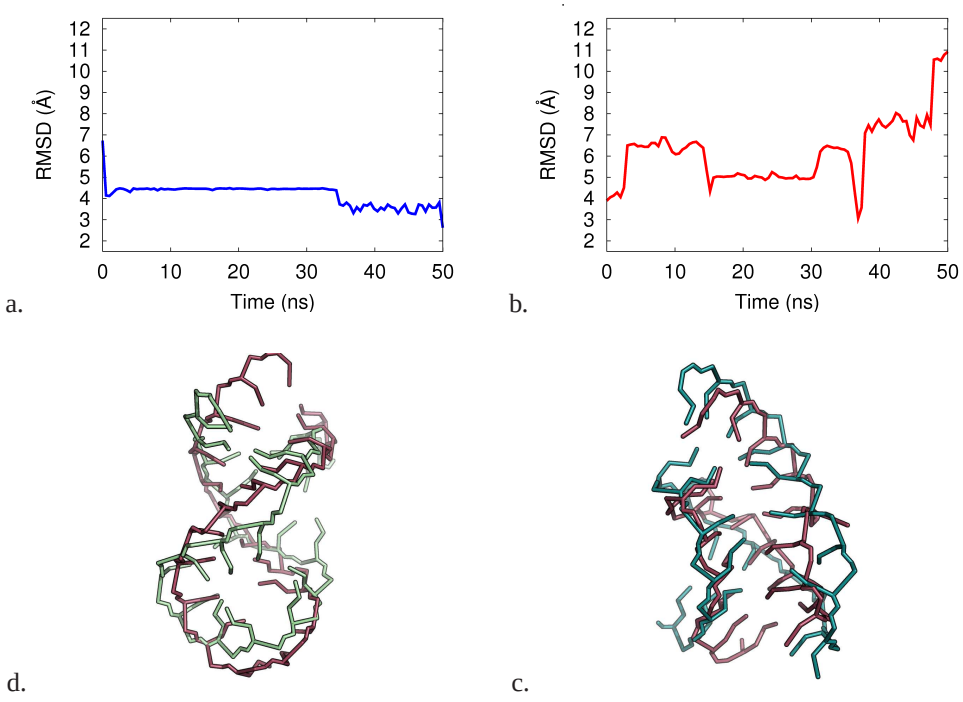


Figure 2. MD-evolution of the RMSD with respect to the NMR structure of 1EOR (a) and 2G1W (b). The RMSD is calculated on the P, O5', C5', C4' and C1' particles using the nucleotides 2-21 (the first and last are flexible by NMR). (c) 2G1W : superposition of the NMR structure (pink) on the MD-structure found at 36 ns (blue). This structure has 3 among 7 native H-bonds. (d) 1EOR : superposition of the NMR structure (pink) on the center of the second cluster predicted by REMD (green).

References

1. Sorin , E.J.,Bradley,J.,Nakatami, J., Rhee, Y. M., Jayachandran, G., Vishal, S., Pande, V.S. *Does native state topology determine the RNA folding mechanism?* J. Mol. Biol. **337**, 789-797, 2004.
2. Knotts, T. A., Rathoe, N., Schwartz, D. C., de Pablo, J. J. *A coarse grain model for DNA.* J. Chem. Phys. **126**, 2007.
3. Hyeon, C. Thirumalai, D. *Chemical theory and computation special feature: mechanical unfolding of RNA hairpins.* Proc Nucl. Acad. Sc. USA. **102**, 6789-6794, 2007.
4. Ding, F., Sharama, S., Chalasani, P., Demidov, V. V., Broude N. E., Dokholyan, N. V. *Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms.* RNA **14**, 1-10, 2008.
5. Derreumaux, P., Mousseau, N. *Coarse grained protein molecular dynamics similations.* J. Chem. Phys. **126**, 025101, 2007.
6. Maupetit, J., Tuffery, P., Derreumaux, P. *A coarse grained protein force field for folding and structure prediction.* Proteins **69**, 394-408, 2007.
7. Derreumaux, P. *Simulating the early steps of amyloid fibril formation and disassembly.* See accompagning paper in this volume.

Understanding of High Pathogenicity of the Avian Influenza Virus H5N1: Why H5 is Better Cleaved by Furin

Panita Decha[1], **Peter Wolschann**[2], **and Supot Hannongbua**[1]

[1] Department of Chemistry, Faculty of Science, Chulalongkorn University, Phayathai Road, Patumwan, Bangkok 10330, Thailand
E-mail: supot.h@chula.ac.th

[2] Institute of Theoretical Chemistry, University of Vienna, Waehringer Strasse 17, A-1090, Vienna, Austria
E-mail: karl.peter.wolschann@univie.ac.at

The origin of high pathogenicity of the emerging avian influenza H5N1 due to the -RRRKK- insertion at the cleavage loop of the hemagglutinin H5 was studied using molecular dynamics technique in comparison with the non-inserted H5 and H3 bound with furin active site. The cleavage loop of the highly pathogenic H5 was found to bind strongly to the furin cavity, serving as a conformation suitable for the proteolytic reaction. Experimentally, the -RRRKK- insertion was also found to increase the cleavage of hemagglutinin by furin. The simulated data provide a clear answer to the question why inserted H5 is better cleaved by furin than the other subtypes, explaining the high pathogenicity of avian influenza H5N1.

1 Introduction

Proteolytic activation of the hemagglutinin (HA) is essential for viral infectivity and for spread of the avian influenza virus. This process is determined by a cleavage reaction at HA's cleavage site, a conserved arginine, by host proteases. Insertion of the -RRRKK- residues into the low pathogenic avian influenza (LPAI) cleavage site is known to potentially activate infectivity of viruses, i.e., the LPAI viruses, which then become high pathogenic avian influenza (HPAI) viruses, allowing highly virulent strains to be cleaved by furin, an ubiquitously expressed protease.The proposed cleavage mechanism is shown in Fig. 1. Understanding of this fact, why furin cleaves the inserted hemagglutinin strains better than non-inserted strains, is the goal of this study. Therefore, molecular dynamics simulations were carried out for the three complexes, HPH5-FR, LPH5-FR and LPH3-FR. The investigation was focused to intra- and intermolecular interactions and geometries of the substrate-furin complex, potentially involved in the cleavage mechanism.

2 Methodology

The initial model for the HPH5-FR loop (RERRRKKRGL) was built up using the sequence alignments and the atomic coordinates of the X-ray structure (residues 322-331: NVPEKQTQGL) of the HA0 of H3 and dec-RVKR-cmk inhibitor of furin[1] as a template, performed by using the homology module of the Insight II program. For LPAI subtype H5, the initial structure of the cleavage loop (NVPQRETRGL) was constructed using the backbone atoms of the HPH5 loop built previously. The HA$'$s cleavage loop complexed

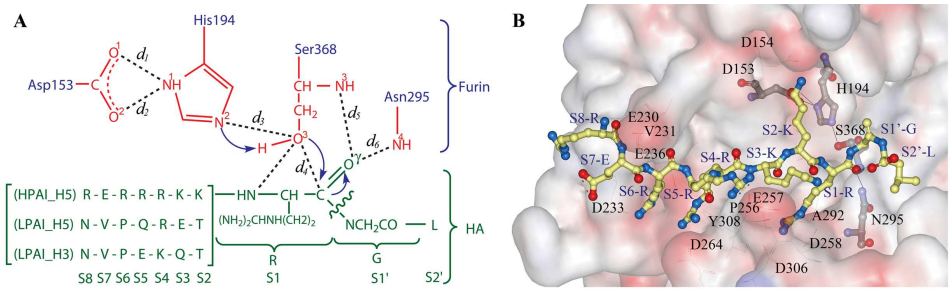


Figure 1. **(A)** Proposed cleavage mechanism of HA by furin and definitions of d_1-d_6. **(B)** Loop of HPAI H5 (ball and stick model) in the electrostatic surface of furin. Blue and red represent positively and negatively charged amino acid residues, respectively.

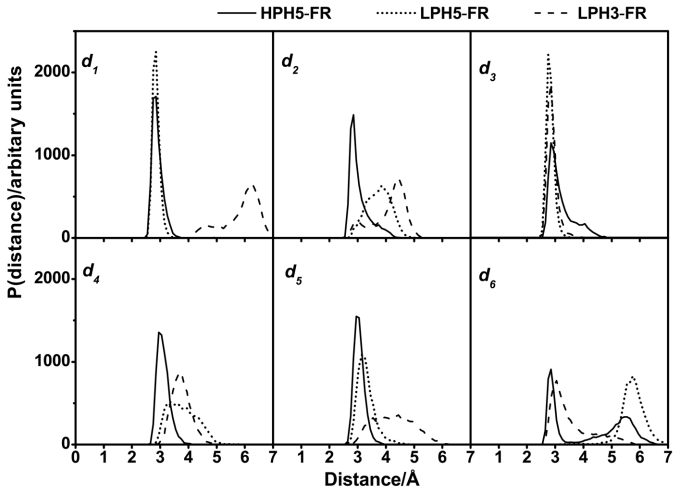


Figure 2. Distributions of the d_1-d_6 distances defined in Fig. 1 for the three simulated systems, sampling from 0.75 to 2.0 ns in MD simulations.

with furin, generated by all HA heavy atoms of S1-S4, were superimposed with the crystal structure of the dec-RVKR-cmk inhibitor while the HA backbone atoms of S5-S8 and $S1' - S2'$ were superimposed with the crystal structure of the HA0 loop of H3. MD simulations for the HA's cleavage loop complexed with furin, HPH5-FR, LPH5-FR and LPH3-FR, were carried out using the SANDER module of AMBER 7.

3 Results and Discussion

To search for detailed information on molecular level, selected structural parameters (d_1-d_6 defined in Fig. 1) were plotted in Fig. 2. The structures of the complexes are in detail described in three regions, the arrangement of the catalytic triad, the attachment of the catalytic Ser368 to the reactive S1-Arg, and the formation of the oxyanion hole. From

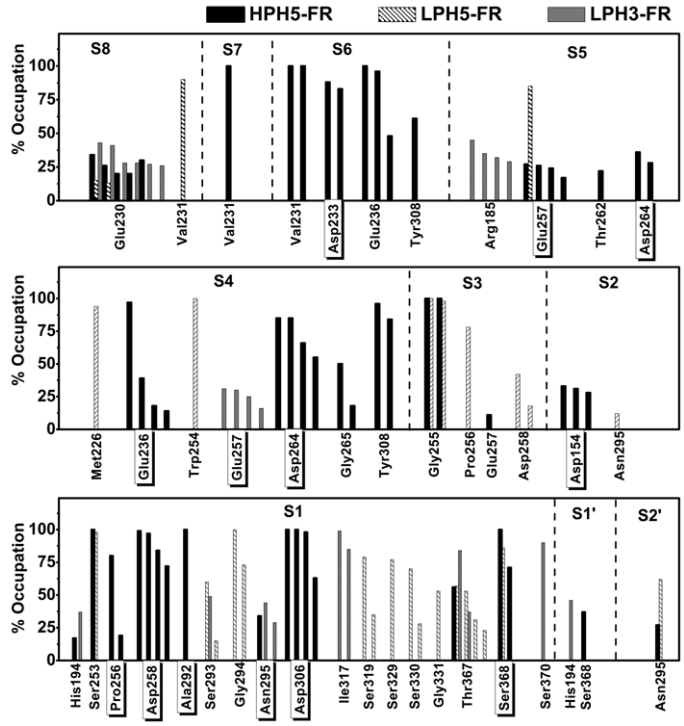


Figure 3. Percent occupations of hydrogen bonds between furin and the ten HA residues where the residues with a box around the label represent experimentally detected bonds for the inhibitor-furin complex.

the result, a sharp peak was found only in the HPH5-FR complex and occurred at suitable distances, signifying the rigidity of the complex which thus serves as a more appropriate configuration for the nucleophilic attack. In contrast, for the other two complexes the peaks show a broad distribution and occure at significantly larger distances.

To assess the stability of the hemagglutinin loop binding into furin protease, percentage and number of hydrogen bonds between each HA residue and the active site residues of the target enzyme, furin, were evaluated and plotted in Fig. 3. Considering the role of the -RRRKK- insertion more hydrogen bonds and a higher percentage occupation between the S2-S6 residues of HA and the surrounding residues of furin were found for HPH5-FR in comparison with the two LPAI systems. This means that the -RRRKK- insertion can directly help to hold the substrate in place.

4 Conclusion

In conclusion, the -RRRKK- insertion in the HPH5-FR, in particular the two arginines at S4 and S6 positions helps directly to hold the HA's cleavage loop in place by forming strong hydrogen bonds between residues of HA and furin. This consequently leads to an active conformation of the HPH5-FR complex suitable for the acylation reaction and is the primary source of high pathogenicity of the avian influenza viruses subtype H5N1.

Acknowledgments

P.D. would like to thank the Sandwich Ph.D. Program from the Commission on Higher Education. The Computational Chemistry Unit Cell, Faculty of Science, Chulalongkorn University and Institute of Theoretical Chemistry, University of Vienna provided research facilities, software packages and computing times.

References

1. S. Henrich, A. Cameron, G. P. Bourenkov, R. Kiefersauer, R. Huber, I. Lindberg, W. Bode, and M. E. Than. *The crystal structure of the proprotein processing proteinase furin explains its stringent specificity*, Nature Struct. Biol. **10**, 520–526, 2003.

Bound Water as a Tool to Detect Soluble Amyloid Oligomers and Amyloid Protofibrils, the Early Stage of Development of the Alzheimer's Disease

Florin Despa

Department of Surgery, The University of Chicago, Chicago, IL 60637, USA
E-mail: fdespa@uchicago.edu

John von Neumann Institute for Computing, Research Centre Jülich, 52425 Jülich, Germany

We investigated the fundamental molecular-level behavior of water molecules adjacent or very close to the surface of amyloid peptides, soluble oligomers and protofibrils. The foucus was on prospecting a reliable means to predict, measure and interpret the magnetic behavior of such water molecules, especially to differentiate their magnetic behavior from that of other, more bulk-like molecules. The long term goal of this will be developing a diagnostic method to enable magnetic resonance relaxation measurements to reveal the nature of the surfaces to which those water molecules are associated, and hence to provide a probe of the extent of tissue damage in the early stage of development of conformational diseases, such as the Alzheimer's disease.

1 Introduction

At structural level, proteins display both binding (hydrophilic) and unbinding (hydrophobic) sites for water molecules. Most of hydrophobic moieties are buried inside the structure when proteins are in native states. They become exposed when proteins unfold (or misfold), which weakens the effective interaction of surrounding water with the protein surface[1]. Recent studies[2] suggest that isomers bearing pathological defects, i.e. proteins exhibiting poorly dehydrated backbone hydrogen bonds (dehydrons)[3,4], are characterized by an average energy of hydration that is also significantly bellow that corresponding to native proteins. Though these structural defects generally interact with nearby water, the entropy of this water is unexpectedly large and the residence time much shorter than at hydrophilic sites[2]. Thus, both unfolded/misfolded and pathogenic proteins exhibit an increased number of surface patches where the local water is less structured and has an increased mobility[5,6], reminiscent of the bulk solvent. Moreover, energetic considerations suggest that isomers with considerable bulk-like hydration tend to aggregate[7]. Here, we summarize the results of our initial studies which suggest that different morphological states of aggregated isomers differ by hydration distribution profiles and water magnetic resonance (MR) signals. The results help explain the MR contrast patterns of amyloids, a subject of long controversy, and suggest a new approach for identifying unusual protein aggregation related to disease.

Extracellular amyloid β (Abeta) deposits are prominent and universal Alzheimer's disease (AD) features but plaque abundance does not reflect the actual degree of the neuronal injury in AD patients[8]. The incipient assemblies formed by Abeta peptides inside the cell, such as soluble oligomers and fibrillar tangles, also have potent neurotoxic activities, and in fact, may be the proximate effectors of the neuronal injury and death occurring in AD[8].

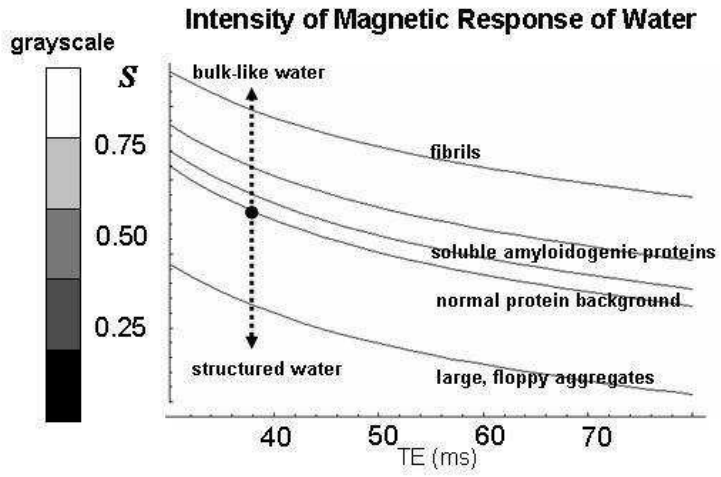


Figure 1. The relative change in pixel intensity corresponding to distributions of constrained water in local environments containing normal protein background, soluble amyloidogenic anssemblies, protofibrils and large protein aggregates.

Thus, the most visible manifestation of amyloid accumulation, i.e. extracellular Abeta plaques, could actually be a defense mechanism employed to avoid serious tissue degradation while the major toxic effects to the gray and white matter neurons are mediated by soluble Abeta composites and protofibrils. Therefore, developing strategies to detect in vivo the Abeta structures at an early stage are crucial for treating, inhibiting the progression of disease and preventing some of the devastating effects of AD.

2 Methods

Based on the scaled particle theory[9], we derived a mathematical algorithm that describes the recompartmentalization of water following the addition of test isomers in the local environment and the association of these isomers in composites of given structures (i.e. oligomers, protofibrils and large protein aggregates). The mathematical equations for the recompartmentalization of water are given in terms of the local hydration fraction (η) and packing density (ρ) of the newly formed molecular assembly, as described previously[7]. Then, we use standard relaxometry equations[7] to correlate the intensity (S) of the magnetic relaxation response of water with η and ρ for each structural archetype (single isomers, oligomers, protofibrils and large aggregates). Based on this model, we investigated the relationship between the intensity of the magnetic response, the amount of constrained water in the local volume and morphological characteristics of the protein system. The predicted magnetic signal intensity for these systems is compared to that corresponding to native protein background (control).

3 Results and Discussion

The present study suggests that a cellular environment containing pathogenic proteins and distinct morphological structures formed by the association of these proteins differ from normal protein background by the quantity of water structured at interfaces. The reorganization of hydration water molecules leads implicitly to a redistribution of water molecules among the characteristic clusters of the relaxation times[2,6]. Normally, such differences lead to distinct MRI signals. The fact can be inferred from Fig. 1 where we show generic curves of the relative change in pixel intensity corresponding to the magnetic response of constrained water in various local environments (i.e. normal protein background, test isomers, oligomers, protofibrils, large protein aggreagtes). We can notice that the hyper-intensity curve corresponding to compact aggregates is the closest to the pixel intensity of bulk water. Therefore, regions containing small compact aggregates are likely to show brighter than the normal protein gray background on MR images. On the contrary, the formation of large, floppy protein aggregates by including large amounts of test isomers and caged water, leads to the occurrence of hypo-intensity on the characteristic MR images. Hence the MR images of such microscopic environments will show darker on the grayscale. The characteristic magnetic sensitivity for a system containing additional test isomers in a morphological state as described in above is represented in Fig. 1 by the bottom curve. We must note that this hypo-intensity signal is apart from the contrast produced by the presence of iron in plaques[10–14]. The present study shed light on a current controversy, namely that amyloids can display both dark and bright spots when compared to the normal, gray background tissue on MR images[10,11]. In addition, our findings suggest that the bright spots are more likely to correspond to amyloids in their early stage of development.

4 Concluding Remarks

The present results are in support of existing experimental data[10,13,14] showing that the presence of iron in amyloid plaques is not always decisive for detecting the AD. We have presented compelling theoretical evidence that placing additional constraints on water dynamics (i.e., caging water) and/or redistribution of constrained water fractions also play a significant role in correlating the intensity of the MR signal with the amyloid load. Our results help to better understand various biophysical mechanisms that set the MR signal of water surrounding amyloidogenic proteins and their model aggregates. Our study may prove useful in generating new testable statements on circumstances related to the presence of amyloidogenic proteins in a given aqueous environment. Progress in understanding the chemistry effects induced by such molecular entities on dynamics of surrounding water in combination with data from new MR spectroscopic methods for determining the over-expression of abnormal proteins and their state of association in the cell can help designing efficient MR imaging protocols to be used in detecting the early molecular alterations in amyloidogenic diseases.

Acknowledgments

As August 1, 2008, the author is with the Dept. of Pharmacology at the University of California. The author thanks RS Berry, A. Fernandez, R. Scott and U.H.E. Hansmann for useful discussion.

References

1. V.P. Denisov and B. Halle, *Hydration of denatured and molten globule proteins*, Nat. Struct. Biol **6**, 253–260, 1999.
2. A. De Simone, G.G. Dodson, C.S Verma, A. Zagari and F. Fraternali, *Prion and water: Thight and dynamical hydration sites have a key role in structural stability*, Proc. Natl. Acad. Sci. USA **102**, 7535–7540, 2005.
3. A. Fernandez and R.S. Berry, *Extent of hydrogen-bond protection in folded proteins: A constraint on packing architectures*, Biophys. J. **83**, 2475–2481, 2002.
4. A. Fernandez, *What factor drives the fibrillogenic association of beta-sheets*, FEBS Lett. **579**, 6635–6640, 2005.
5. F. Despa, A. Fernandez and R.S. Berry, *Dielectric modulation of biological water*, Phys. Rev. Lett. **93**, 228104, 2004.
6. F. Despa, *Biological water, its vital role in macromolecular structure and function*, Ann. N.Y. Acad. Sci. **1066**, 1–11, 2005.
7. F. Despa and R.S. Berry, *The origin of long range attraction between hydrophobes in water*, Biophys. J. **92**, 373–378, 2007.
8. D. Watson, E. Castano, T.A. Kokjohn, Y.M. Kuo, Y. Lyubchenko, D. Pinsky, E.S. Jr. Connolly, C. Esh, D.C. Luehrs, W.B. Stine, L.M. Rowse, M.R. Emmerling and A.E. Roher, *Physicochemical characteristics of soluble oligomeric Abeta and their pathologic role in Alzheimer's disease*, Neurol. Res. **27**, 869-881, 2005.
9. H. Reiss, *Scaled particle methods in the statistical thermodynamics of fluids*, Advan. Chem. Phys. **9**, 1–84, 1965.
10. M. Dhenain, N. Privat, C. Duyckaerts and R.E. Jacobs, *Senile plaques do not induce susceptibility effects in T2*-weighted MR microscopic images*, NMR Biomed **15**, 197–203, 2002.
11. S.P. Lee, M.F. Falangola, R.A. Nixon, K. Duff, and J.A. Helpern, *Visualization of beta-Amyloid plaques in a transgenic mouse model of Alzheimer's disease using MR microscopy without contrast reagents*, Magn. Reson. Med. **52**, 538–544, 2004.
12. C.R. Jr. Jack, T.M. Wengenack, D.A. Reyes, M. Garwood, G.L. Curran, B.J. Borowski, J. Lin, G.M. Preboske, S.S. Holasek, G. Adriany and J.F. Poduslo, *In vivo magnetic resonance microimaging of individual amyloid plaques in Alzheimer's transgenic mice*, J. Neurosci. **25**, 10041–10048, 2005.
13. N. El Tannir El Tayara, A. Volk, M. Dhenain and B. Delatour, *Transverse relaxation time reflects brain amyloidosis in young APP/PS1 transgenic mice*, Magn. Reson. Med. **58**, 179–184, 2007.
14. M.F. Falangola, V.V. Dyakin, S.P. Lee, A. Bogart, J.S. Babb, K. Duff, R. Nixon, and J.A. Helpern, *Quantitative MRI reveals aging-associated T2 changes in mouse models of Alzheimers disease*, NMR Biomed **20**, 343–351, 2007.

Stabilizing Regions in Membrane Proteins

**Frank Dressel, Annalisa Marsico, Anne Tuukkanen,
Rainer Winnenburg, Michael Schroeder, and Dirk Labudde**

Biotechnologisches Zentrum , TU Dresden,
Tatzberg 47/49, 01307 Dresden, Germany
E-mail: firstname.lastname@biotec.tu-dresden.de

Around one third of a typical genome consists of membrane proteins. Misfolding of membrane proteins can often be linked to diseases, so that it is of great importance to understand, which residues and interactions are crucial for its stability. We developed a coarse-grained model to predict stabilizing regions in membrane proteins. We compare the model to experimental data from Single Molecule Force Spectroscopy (SMFS) and literature to evaluate the effects of mutations on function and stability of five membrane proteins. The aim of this study is to describe all these data in an unified context, the interaction energies of amino acids in a coarse grained model to gain a deeper insight into membrane proteins.

1 Introduction

Integral membrane proteins account for about 20%-30% of the open reading frames in a typical genome but, despite their central importance for all organisms, the number of known structures remains small giving raise to the need for sequence based methods.

Single Molecule Force Spectroscopy (SMFS) allows detecting and locating interactions stabilizing a membrane protein. During continuous stretching of the molecule, the applied forces and the extension are measured. The resultant force-distance curve reflects subsequent unfolding events[1], the so called force peaks. The so-called Worm-like-chain model (WLC)[2] relates the position of these unfolding barriers to positions in the primary sequence of the protein. Besides the SMFS data, there are a lot of mutation experiments in the literature, where the effect of a certain mutation on stability and function of a membrane protein is discussed.

In this work, we developed a coarse grained model, which is able to describe the above mentioned experimental results. To the best of our knowledge, this is the first time, that such a simplified model is used for predicting correctly unfolding barriers and estimating the influence of mutations based on sequence.

2 Methods

For the representation of a protein, we constrained ourself to the C_α/C_β atoms of each amino acid. To estimate the energy of the interaction between amino acids, we defined a solvation energy e_i based on the probability distribution for each amino acid in an membrane protein to be inside (solved) or outside of the protein[3]. Inside means facing other amino acids, outside means facing the phospholipids. The combination of two solvation energies results in a contact energy for an interaction between two spatial neighbored amino acids.

Based on the tertiary structure we estimate the energy E_i with which each amino acid i interacts with the rest of the protein by summing up the interaction energies between i and all other amino acids. Two amino acids interact if they are not direct neighbors in sequence and if their C_β atoms are less than $8\mathring{A}$ apart. This gives an energy profile containing the energy E_i for each amino acid in the protein. A deep valley in this profile gives raise to an energy barrier in the model.

To estimate the influence of a mutation based on the sequence only, we calculate a so called solvation energy profile, which contains for each amino acid i the solvation energy e_i. This profile is, in contrast to the energy profile, independent from the tertiary structure of a protein.

3 Results

For our analysis, we used the following five membrane proteins: Bacteriorhodopsin, Halorhodopsin, Bovine rhodopsin, Na^+/H^+ antiporter and Aquaporin-1. Up to now, these are the only proteins, for which SMFS data are available.

3.1 Prediction of Unfolding Barriers

For the five proteins mentioned above we predicted the barriers based on their energy profiles. For the comparison with the experiment, we considered the published barrier position from the experiment plus an experimental error of $\Delta_{exp} = \pm 4$ amino acids. This is due to systematic errors in the experiment like intrinsic movement of the cantilever. These errors are reported to be in the range of 3-7 amino acids[4], depending on the used cantilever.

On average 61% of the barriers are detected by our method. A match between a predicted barrier and an experimentally determined one is established if their boundaries are within a certain distance.

A remarkable result is the following: all unfolding barriers for all proteins that, according to SMFS measurements, are found having 100% of probability of occurrence (main peaks) are correctly predicted by our method. This suggests that major unfolding barriers in the experiments are due to energetic reasons. Minor events are more difficult to detect with this approach.

Energy profiles can also be used for detecting smaller stabilizing regions in a protein. The helix C in bacteriorhodopsin for example is known for its important part in the proton pumping pathway[5]. The mutations T90A and T90V in helix C for example decrease the activity down to 10% and 20% compared to the wild type[6]. This is thought to be partly due to the missing stabilizing interaction of the threonine. Interestingly, alanine and valine are both more repulsive in our model than threonine itself. These mutations have a destabilizing effect on this region in helix C.

3.2 Predict Effects of Mutations Based on the Sequence

To quantify the ability of the model to predict the influence of mutations based on sequence alone, we screened the literature for single point mutations, which influence stability or function of the used membrane proteins. For this, we applied a text mining approach,

which searches in PubMed abstracts. We only used the publications, where a single point mutation was discussed in the context of stability or stability related function. If an appropriate mutation was found in the literature, we compared the solvation energies of both residues to decide, if the mutation was stabilizing, slightly stabilizing, slightly destabilizing and destabilizing. 25 out of 35 mutational effects reported in the literature correlate with the predictions from the energy profiles and thus can be related to stability issues in terms of amino acid interactions.

4 Summary

We used a coarse grained model to assess the influence of interactions between amino acids in membrane proteins. It is based on the probability distribution for an amino acid to be inside or outside of an membrane protein. We found that major SMFS unfolding barriers are related to strongly interacting amino acids, whereas minor barriers are related to non-energetic reasons. The activity of bacteriorhodopsin and the effect of mutations at site Thr90 is related to an increase of the interaction energy of this amino acid if mutated to Ala or Val. Although these two use cases require knowledge about the tertiary structure of a protein, it is possible to extend the scope of the model to a purely sequence based method. Compared with a list compiled from literature, 25 out of 35 mutations can be related to energetic reasons in the model by this sequence-only approach. Concluding, the proposed coarse grained interaction scheme can be applied successfully to a wide range of applications in the context of membrane proteins.

References

1. F Oesterhelt, D Oesterhelt, M Pfeiffer, A Engel, HE Gaub, and DJ Muller, *Unfolding pathways of individual bacteriorhodopsins.*, Science, **288**, no. 5463, 143–6, Apr 2000.
2. C Bustamante, J F Marko, E D Siggia, and S Smith, *Entropic elasticity of lambda-phage DNA*, Science, **265**, no. 5178, 1599–1600, Sep 1994, Letter.
3. F Dressel, A Marsico, A Tuukkanen, R Winnenburg, M Schroeder, and D Labudde, *Analyzing stable regions in membrane proteins using a coarse grained energy model*, submitted, 2008.
4. A Kedrov, C Ziegler, H Janovjak, W Kuehlbrandt, and DJ Muller, *Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy.*, J Mol Biol, **340**, no. 5, 1143–52, Jul 2004.
5. H Luecke, HT Richter, and JK Lanyi, *Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution.*, Science, **280**, no. 5371, 1934–7, Jun 1998.
6. A Peralvarez-Marin, M Marquez, JL Bourdelande, E Querol, and E Padros, *Thr-90 plays a vital role in the structure and function of bacteriorhodopsin.*, J Biol Chem, **279**, no. 16, 16403–9, Apr 2004.

Loop Parameterization and RNA Secondary Structure Folding

Thomas R. Einert[1], **Paul Näger**[1], **Henri Orland**[2], **and Roland R. Netz**[1]

[1] Physik Department, Technische Universität München,
85748 Garching b. München, Germany
E-mail: {einert, netz}@ph.tum.de

[2] Institute de Physique Théorique, CEA Saclay,
91191 Gif-sur-Yvette Cedex, France
E-mail: Henri.Orland@cea.fr

Loops are abundant in native RNA structures and proliferate close to the unfolding transition. By including a statistical weight $\sim \ell^{-c}$ for loops of length ℓ in the recursion relation for the partition function, we show that the calculated heat capacity depends sensitively on the presence and value of the exponent c, even of short tRNA. For homo-RNA we analytically calculate the critical temperature and critical exponents which exhibit a non-universal dependence on c.

We calculate the partition function of the RNA secondary structure using a formulation that allows to accurately include the statistics of terminal, internal, as well as multi-loops.[1] The statistical weight of a secondary structure depends on the free energy of base pair formation, which has been determined experimentally[2], but also on the entropy loss of loop formation. Polymer theory predicts the configurational weight of a loop consisting of ℓ bases to decay as ℓ^{-c} where the exponent c is universal. The loop exponent is $c_{\rm ideal} = 3/2$ for an ideal polymer and $c_{\rm SAW} = d\nu \simeq 1.76$ for an isolated self avoiding loop. However, helices which emerge from the loop decrease the number of configurations and consequently increase c even further.[3] For instance, one obtains $c_1 = 2.06$, $c_4 = 2.16$ for the two types of loops which appear in the native structure of yeast tRNA-phe, see Fig. 1b.

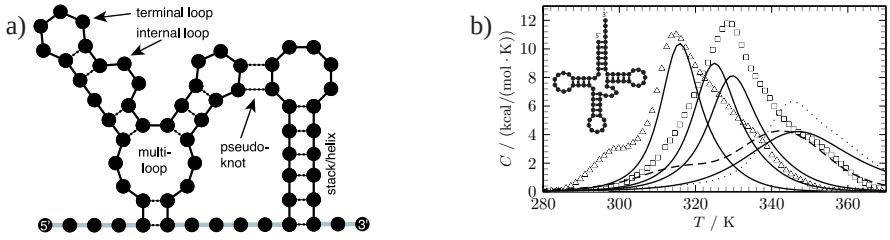


Figure 1. a) Schematic representation of a secondary RNA structure. Solid lines denote the RNA backbone, broken lines base pairs, and gray lines non-nested backbone bonds that are counted by the variable M; here $M = 11$. b) Experimental heat capacity of the tRNA-phe of yeast for NaCl concentrations $20\,\mathrm{mM}$ (triangles) and $150\,\mathrm{mM}$ (squares).[4] Solid lines show results using Eq. (1) with loop exponents $c = 3.0,\ 2.16,\ 1.76,\ 0$ (from left to right), compared with the results from the `Vienna package`[5] which uses a linearized multi-loop entropy (dashed curve). The dotted curve is obtained with $c = 3$ and the same energy parameter set as for the solid curves, except for the loop initiation penalty which was omitted. The inset sketches the low-temperature secondary RNA structure obtained from Eq. (1) in agreement with experimental crystal-structure studies.

A valid secondary structure is a list of all base pairs, where pseudo-knots[6] are not allowed, *i. e.* for any two base pairs (i, j) and (k, l) with $i < j$, $k < l$, and $i < k$ we have either $i < k < l < j$ or $i < j < k < l$. In our notation, the canonical partition function $Q_{i,j}^M$ of a sub-strand from base i at the 5' end through j at the 3' end depends on the non-nested backbone-length M,[7,8] see Fig. 1a. The recursion relations for the partition function read then

$$Q_{i,j+1}^{M+1} = Q_{i,j}^M + \sum_{k=i+M+1}^{j-N_{\rm loop}} Q_{i,k-1}^M Q_{k,j+1}^0 \quad \text{and} \tag{1a}$$

$$Q_{k,j+1}^0 = \sum_{h=1}^{(j-k-N_{\rm loop})/2} w_{(k+h,j+1-h)}^{(k,j+1)} \sum_{m=1}^{j-k-1-2h} \frac{Q_{k+1+h,j-h}^m}{(m+2)^c} \; . \tag{1b}$$

Eq. (1a) describes elongation of an RNA structure by either adding an unpaired base (first term) or by adding an arbitrary sub-strand $Q_{k,j+1}^0$ that is terminated by a helix. Eq. (1b) constructs $Q_{k,j+1}^0$ by closing structures with m non-nested bonds, summed up in $Q_{k+1+h,j-h}^m$, by a helix of length h, which is weighted with a sequence dependent Boltzmann factor w. $N_{\rm loop} = 3$ is the minimum number of bases in a terminal loop. The unrestricted partition function of the entire RNA is given by $Z_N = \sum_M Q_{0,N}^M$. We implement the recursion relation, Eq. (1), numerically using a free energy parameter set.[2] In Fig. 1b we show the experimental heat capacity of the tRNA-phe of yeast compared with our predictions from Eq. (1) using $C = T\partial^2(\mathrm{k_B}T \ln Z_N)/\partial T^2$. The heat capacity peak corresponds to the gradual melting of the secondary structure. Although the RNA consists of just 76 nucleotides and is therefore far from the thermodynamic limit where one expects asymptotic effects to be important, the loop exponent c has drastic effects. Increasing c from $c = 0$ to $c = 3$ destabilizes the structure and decreases the melting temperature by more than 30 K (solid lines). It is difficult to directly compare experimental and theoretical curves as the energy parameters were determined at 1 M NaCl concentration[2], while experimental data is only available at 20 mM and 150 mM. Current implementations of secondary structure prediction or partition function calculation approximate the entropy for multi-loops by an affine function $\ln(y^M M^{-c}) \approx \delta_0 + \delta_1 M$.[5,9] This in principle corresponds to the usage of the loop exponent $c = 0$, as is corroborated by the near agreement of the results from the Vienna package[5] (broken line) with the results from Eq. (1) using $c = 0$. Most strikingly, the melting temperature as well as the width and the height of the peaks depend dramatically on the loop exponent. Similar studies have been carried out for DNA[10] where the loop exponent has a much weaker effect on denaturation curves.

In a second step, we now consider homo-polymeric RNA, which can be modeled experimentally by using a synthetic sequence like AUAUAU... The goal is to extract the critical asymptotic behavior embodied in Eq. (1) in the thermodynamic limit. We simply give a statistical weight $w = \exp[-\varepsilon/(\mathrm{k_B}T)]$ to each base pair. This renders the system translationally invariant and allows to write $Q_{i,j}^M$ as Q_N^M with $N = j - i$ being the total number of backbone segments of the sub-strand ranging from i through j. This can be viewed as a coarse-graining approximation for natural or random RNA above the glass transition. To proceed, we switch to the grand canonical ensemble where we are able to study the

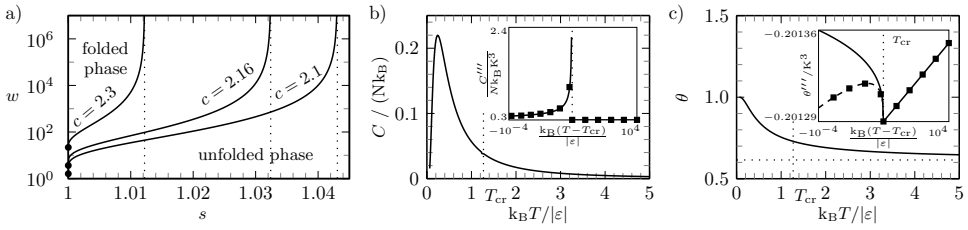


Figure 2. a) Phase diagram for three different values of the loop exponent c as a function of the base pairing weight w and force fugacity s. The dots denote the unfolding transition in the absence of external force, *i. e.* $s = 1$, which is considered in b) and c): Temperature dependence of the b) specific heat C and c) fraction of bound bases θ for $c = 2.3$. The insets show the third derivatives $C''' = \mathrm{d}^3 C/\mathrm{d}T^3$ and $\theta''' = \mathrm{d}^3\theta/\mathrm{d}T^3$ which clearly exhibit singular behavior. Squares denote numerical evaluation of Eq. (2). The leading (solid lines) and next-leading (dashed line) order of the expansion around T_{cr} are shown, according to which C''' diverges with the exponent $\chi = 2/3$ for $c = 2.3$ and θ''' is characterized by the exponent $\lambda = 1/3$.

thermodynamic limit. The grand canonical partition can be calculated exactly

$$\mathcal{Z}(z,s) = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty} z^N s^M Q_N^M = \frac{\kappa(z)}{1 - sz\kappa(z)}\ , \tag{2}$$

where $\kappa(z)$ is determined by the equation $\kappa = 1 + w/\kappa \mathrm{Li}_c(z\kappa)$. $\mathrm{Li}_c(x) = \sum_{n=1}^{\infty} x^n/n^c$ is the polylogarithm. The force fugacity s is $s = 1$ if no force is applied to the ends and $s > 1$ if the molecule is stretched. In Fig. 2a we show the resulting phase diagram of RNA in terms of w and s for different values of the loop exponent c. We observe a very weak phase transition only if $2 < c < c^*$ with $c^* \simeq 2.479$. For $c < 2$, the RNA is always in the folded state, whereas for $c > c^*$ RNA is always unfolded, irrespective of the temperature. The critical exponents as well as the order, which is at least four, of the phase transition depend on the loop exponent c and are calculated exactly.

The conclusion is that while the dependence of critical properties on the loop exponent c is experimentally and numerically[11] difficult to access and therefore largely irrelevant, the dependence of non-critical properties on c is important.

Acknowledgments

Support from the *Elitenetzwerk Bayern* within the framework of *CompInt* is gratefully acknowledged.

References

1. T. R. Einert, P. Näger, H. Orland, and R. R. Netz, *submitted to Phys. Rev. Lett.* (2008).
2. D. H. Mathews *et al.*, J. Mol. Biol. **288**, 911 (1999).
3. Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
4. P. L. Privalov and V. V. Filimonov, J. Mol. Biol. **122**, 447 (1978).
5. I. L. Hofacker *et al.*, Mon. Chem. **125**, 167 (1994).
6. H. Orland and A. Zee, Nucl. Phys. B **620**, 452, (2002).

7. R. Bundschuh and U. Gerland, Phys. Rev. Lett. **95**, 208104 (2005).
8. M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002).
9. J. S. McCaskill, Biopolymers **29**, 1105 (1990).
10. R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).
11. M. Baiesi, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **91**, 198102 (2003).

Investigation of Structure of Amphipathic Peptides in Different Environments via Replica Exchange Molecular Dynamics Simulations

Özge Engin and Mehmet Sayar

Center for Computational Biology and Bioinformatics, Koc University, 34450, Istanbul, Turkey
E-mail: {*ozengin, msayar*}*@ku.edu.tr*

We performed Replica Exchange Molecular Dynamics (REMD) simulation on a 24-residue amphipathic peptide in order to investigate the equilibrium conformational distribution in different environments.The structure of the peptide has been designed such that it adopts a β-hairpin structure.Experimental findings show that it adapts to β-hairpin at the air/water interface.It has the amino acid sequence of -KICVRWQYRVQ (D-P) GDICFDVNFDVH- in which D-P and G residues contribute to the formation of a β-turn structure.We observe that the peptide adapts to coil structure in vacuum whereas it mostly adapts to β-hairpin in bulk water.At the air/water interface it oriented itself in such a conformation that the hydrophobic residues are lined towards the air, on the other hand, hydrophilic residues are lined towards water.

1 Introduction

Amphipathic peptides have alternating hydrophobic and hydrophilic residues in a recurring pattern such that the periodicity in this recursion organizes the non-local interactions among the different types of residues, and this organization among the residues leads to the formation of the type of the overall conformation of the peptide.The equilibrium conformational distribution of amphipathic peptides, as well as others, varies in different environments. Understanding of the probable source of difference in the conformational distribution probabilities is important for both the design of novel peptides and nanomaterials with desired features.

2 Methods

REMD simulations[1] for the systems studied were performed with the Gromos 53A6 forcefield[2] implemented in Gromacs (3.3.1) simulation package[3].The SPC216 water model[4]and NVT ensemble were used for all types of simulations. The temperature ranged from 278 K to 320 K in REMD simulations.The exchange probability (computed as the ratio between the successful exchanges and the total number of trials) varied between 10% and 20% for each pair of neighboring replicas; the average exchange probability was approximately 11%, a level ensuring an efficient exploration of the conformational space[1].The water simulation was performed for 80 ns whereas the vacuum simulation was performed for 10 ns.After equilibration was reached, the data was collected and clustered by using the algorithm as described in Daura et. al[5] with the backbone rmsd of 0.15 and 0.20 nm for bulk water and interface simulations, respectively. The secondary structure assignment was made using the STRIDE algorithm[6].

3 Results

3.1 Vacuum Simulations

The simulation was started from an extended structure, and after approximately 1 ns, it collapsed and remained in a coil structure for the remaining part of the simulation. For this reason, the conformations and the corresponding percentage probability values obtained in vacuum simulation are not given here. These results suggest that the peptide need an assistance of water molecules to adapt to proposed β-hairpin.

3.2 Bulk Water Simulations

The peptide adapts mostly to proposed β-hairpin in bulk water indicating the importance of water molecules for the formation of that structure.However, it has also preference for other conformations other than this preferred β-hairpin structure. These conformations emerge as a result of a reduction in the number of intra-molecular hydrogen (H) bonds.These results suggest that water molecules and the number and the residue types in which the H-bonds are made are important for the formation of the β-turn structure. In this structure β-turn is constructed by D-Proline and Glycine residues. The top four conformations are given in Fig. 1.

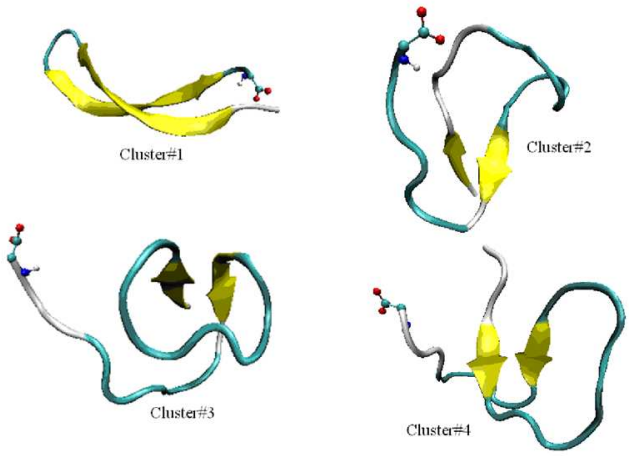


Figure 1. The top four conformations of the peptide obtained upon simulation in bulk water at 300 K. Color codes represent the secondary structure type of each residue. Turn is indicated by green, coil is indicated by white, and sheet structure is indicated by yellow color. The C terminus residues are indicated in CPK representation.

The percentage probability values of these top four conformations are given in Table 1.

Cluster ID	Cluster1	Cluster2	Cluster3	Cluster4
Probability(%)	32.0	8.0	7.84	6.0

Table 1. The percentage probability values of the top four conformations of the peptide in bulk water at 300 K.

3.3 Interface Simulations

The interface simulation was started from a location close to the interface, and the peptide remained there for the rest of the simulation.The top four conformations are given in Fig. 2.

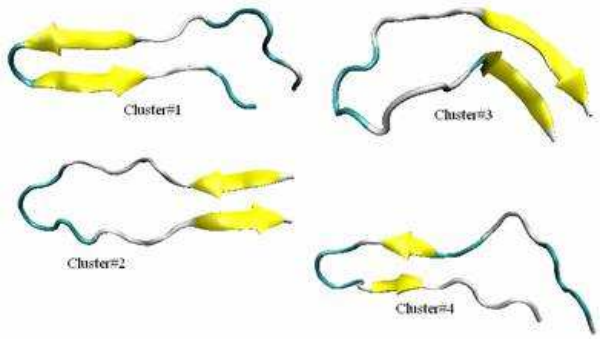


Figure 2. The top four conformations of the peptide obtained upon simulation at the air/water interface at 300 K. Color codes represent the secondary structure type of each residue. Turn is indicated by green, coil is indicated by white, and sheet structure is indicated by yellow color.

The percentage probability values of these top four conformations are given in Table 2.

Cluster ID	Cluster1	Cluster2	Cluster3	Cluster4
Probability(%)	41.0	27.2	7.9	7.7

Table 2. The percentage probability values of the top four conformations of the peptide at the interface at 300 K.

The peptide was aligned parallel to the interface, and it was orientated in such a way that the hydrophobic residues were lined towards the air whereas both the hydrophobic and hydrophilic residues were lined towards water as shown in Fig. 3. However, these are still preliminary results; the equilibrium has not been reached yet.

4 Conclusion and Future Work

We identified the equilibrium conformational distributions of a 24-residue amphipathic peptide via REMD simulations in vacuum, bulk water, and at air/water interface.The peptide adapts to different conformations in different environments. Understanding of the be-

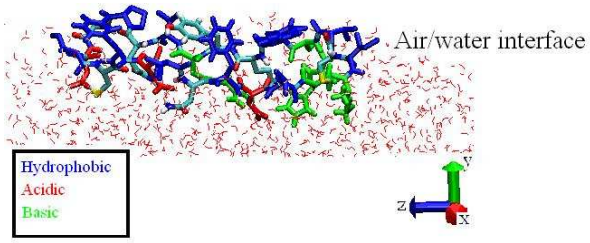


Figure 3. The orientation of residues at the air/water interface. The hydrophobic residues are colored in blue, basic residues are colored in green and the acidic residues are colored in red.

havior of peptides in different environments will help design of novel peptides and nanomaterials with desired features.We are also simulating different amphipathic peptides having different amino acid sequence to investigate the effect of residue specificity on the formation of overall conformations of peptides.Moreover, we are planning to make simulations at interface with larger number of peptides to understand the behavior of both inter and intra molecular interactions.

Acknowledgments

We thank Raymond Tu and his group for scientific discussions, and sharing their experimental results on the peptide molecule. We also thank to Berk Hess and Alessandra Villa for their scientific discussion. This research is supported by grants from TUBITAK (106T575), and Partner Group of Max-Planck Institute for Polymer Research.

References

1. Y.Sugita and Y.Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem.Phys.Lett. **314**, 141, 1999.
2. C. Ooestenbrink,A. Villa,A. Mark and W. F. Gunsteren, *A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation:The GROMOS Force-Field Parameter Sets 53A5 and 53A6*, J. Comp.Chem **25**, 1656, 2004.
3. D. Spoel, A. R. Buuren,E. Meulenhoff,D.P. Tieleman, A. Sijbers,B. Hess, K. Feenstra, E. Lindahl, R. Drunen, and H. Berendsen, *Gromos User Manual Version 3.3.1* (Nijenborgh 4, 9747 AG Groningen, The Netherlands, www.gromacs.org, 2002).
4. H. J.C Berendsen,J. P. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *Molecular Dynamics with Coupling to an External Bath*, J. Chem. Phys. **81**, 3684, 1984.
5. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Peptide folding: When simulation meets experiment*, Angew. Chem. Int. Ed. **18**, 236, 1999.
6. D. Frishman and P. Argos, *Knowledge-based secondary structure assignment*, Proteins: Structure, Function and Genetics **23**, 566, 1995.

Protein-Protein Interaction Prediction

Florian Fink, Stephan Ederer, and Wolfram Gronwald

Institute of Functional Genomics, University of Regensburg, Germany
E-mail: {florian.fink, wolfram.gronwald}@klinik.uni-regensburg.de

Based on amino acid based pair-potentials and intermolecular energies we calculate score distributions for protein-protein complexes that exist in nature and those which do not exist. The distributions of the two groups are then found to be different in maximum and in shape. This opens the possibility to discriminate between complexes that exist and those which do not.

1 Introduction

1.1 From Protein Structure to Complexes

Proteins are an integral component for most of the mechanisms taking part in the cell. One important aspect in the research on proteins is their three-dimensional structure. The most common methods to determine the structure are X-ray cristallography and NMR spectroscopy, and due to them the number of known protein structures is actually rapidly growing. However, cellular functions are rarely carried out by single proteins but by complexes of several interacting proteins. High-throughput methods for detecting protein interactions, like yeast2hybrid, produce a huge number of such expected protein-protein interactions. Unfortunately it is not possible to determine the structures for all of them by experimental methods because there are limitations concerning large or transient complexes. In addition, if possible, the experimental structure determination of complexes is a very time-consuming and challanging process. For that reason computationally approches such as docking algorithms to predict the structure of protein-protein complexes are needed.

1.2 Docking

The hypothesis underlying docking predictions is that the native complex structure is the state with the lowest free energy accessible to the system. There are quite different approches on how to develop docking algorithms but the common, basic idea is to first do a sampling step followed by a scoring step. Scoring means, to analyse the putative complex structures generated in the first step with regard to chemical and physical aspects. Selecting suitable aspects and weighting them in an appropriate way is one of the great challenges in docking. The aim is to rank all putative structures in a way that most of the native-like structures are found in the top part of the ranked output.

2 Motivation

Since protein complexes play a major roll in cellular processes and experimental methods like yeast2hybrid are not always applicable and often contain a considerable number

of false positives[2], there is a need for computational methods predicting protein-protein interactions.

On the other hand, methods providing the three-dimensional structure of known protein complexes (docking algorithms) are already available. In their scoring step a great amount of different possible complex structures of the same two proteins is compared to choose those that are near-native. If this is possible, it must even be possible to do this analysis on complex structures of different protein pairs and by this get information on the probability that two specific proteins do interact at all. That means in other words to do docking with different proteins, even those that do not interact or are not known to do so and finally, after the interpretation of the structures, get as a result whether two proteins are suggested to built complexes in nature or not.

This is actually a computational method to predict protein-protein interaction.

3 Method and Results

3.1 Overview

For becoming able to predict protein-protein interaction what we actually need is a method that discriminates between complexes that exist in nature (native complexes) and those that do not (false complexes). This difference is mesured by (upp to now) three scoring functions (amino acid based pair-potentials, van der Waals energy and electrostatic energy), and becomes apparent in different score distributions for native and false complexes (see figure 1).

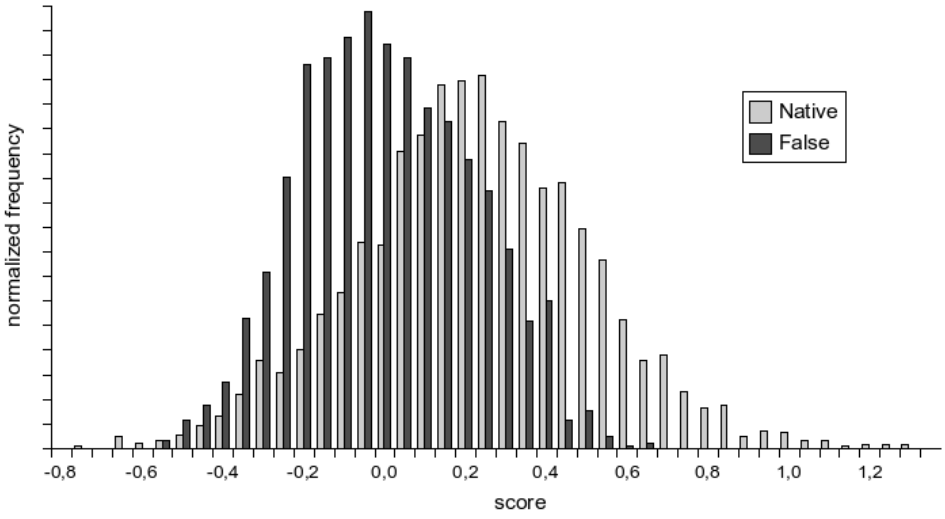


Figure 1. Score distributions from amino acid based pair-potentials.

3.2 Getting Native and False Complexes

Native structures can easily be obtained from the Nussinov database[4] that contains over 2000 non-homologous protein-protein complexes, whereas there does not exist a database for false complexes. For this reason we produced them on our own by somehow abusing the docking program HADDOCK[1]: We docked proteins that are not supposed to build complexes. This ensures that even the false complexes are in the best possible confirmation and hence really comparable to the native ones.

3.3 Scoring Functions

We use **amino acid based pair-potentials** that were obtained by Wolowski et al.[3], and calculated distributions for the native complexes from the Nussinov database and our self-produced false complexes. In figure 1 it can be seen, that scores from the two groups are not totally separated, but that there is an evident difference in the shape of the two curves.

Van der Waals energy and **electrostatic energy** are both calculated between all intermolecular atompairs in the complex and can be combined. Their sum is called interaction energy. We have not yet obtained score distributions for this, but the two examples in table 1 show, that the energies for the native complexes are considerably lower then for the false ones.

Receptor	Ligand	$E_{int} = E_{vdw} + E_{elec}$ [kcal/mol]
Barnase	Barstar	-264.4
Barnase	Soybean trypsin inhibitor	-242.4
Barbase	APPI	-214.0
Barnase	Ovomucoid 3rd domain	-192.0
Barnase	Pancratic secretory trypsin inhibitor	-189.4

Table 1. Intermolecular energies of one nativ complex (shaded in grey) and four false complexes. The energy is always the average of ten complexes that were top ranked from the docking algorithm.

4 Conclusion

We could show that it is possible to find scoring functions that can discriminate between native and false protein-protein complexes. By combination of the three presented scores and maybe even more in future, it will be possible to predict whether a hypothetical complex can be supposed to exist in nature or not.

Acknowledgments

We would like to thank the bavarian genomic network for financial support.

References

1. C. Dominguez, R. Boelens, A. M. Bonvin, *Protein-Protein Docking Approach Based on Biochemical or Biophysical Information*, J. Am. Chem. Soc. **125**, 1731–1737, 2003.
2. C. von Mering, R. Krause, B.Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, *Comparative assessment of large-scale data sets of protein-protein interactions*, Nature **417**, 399–403, 2002.
3. V. Wolowski, *Computational analysis of protein-protein complexes related to knowledge-based predictions of interaction*, (Diploma Thesis, University of Hagen, Germany, 2008).
4. `http://bioinfo3d.cs.tau.ac.il/Interfaces/Non-Redundant`.

Determining RNA Flexibility by Graph Theory: Ribosomal Exit Tunnel as a Case Study

Simone Fulle[1] **and Holger Gohlke**[2]

[1] Department of Biological Sciences, Molecular Bioinformatics Group,
J.W. Goethe-University, Frankfurt, Germany
E-mail: fulle@bioinformatik.uni-frankfurt.de

[2] Department of Mathematics and Natural Sciences, Pharmaceutical Institut,
Christian-Albrechts-University, Kiel, Germany
E-mail: gohlke@pharmazie.uni-kiel.de

We present an analysis of the flexibility characteristics of the ribosomal exit tunnel using concepts grounded in rigidity theory. For this, a new topological network representation of RNA structures had to be developed that allows analyzing RNA flexibility/rigidity based on constraint counting. Applied to the large ribosomal subunit, constraint counting provides new insights in atomic detail into the stability characteristics of the tunnel, which can be linked with the tunnels role in co-translational processes.

1 Introduction

RNA structures are highly flexible biomolecules that can undergo dramatic conformational changes required to fulfill their diverse functional roles. The determination of RNA structures, e.g., by X-ray crystallography provides us with static snapshots along these transitions, whereas the underlying dynamical processes remain largely unclear. To get a more detailed view of the dynamics of biomolecules or to illuminate experimental data, molecular dynamics simulations are very useful and widely applied. Unfortunately, the simulations are still too computationally expensive to investigate large macromolecules like the ribosomal complex on a routine basis. As a much more efficient alternative, concepts from graph theory can be used to determine flexible and rigid regions within a structure.[1] Thereby, a biomolecule is modeled as a topological network, where vertices (joints) represent atoms and edges (struts) represent covalent and non-covalent bond constraints (strong hydrogen bonds, salt bridges, and hydrophobic interactions) as well as angular constraints. Modeling non-covalent constraints appropriately is detrimental to the success of the analysis. Given a network representation, a fast combinatorial algorithm, the *pebble game*,[2] can then be applied to determine the number and spatial distribution of bond-rotational degrees of freedom in the network, which can be related to rigid regions and flexible links in between. Modeling non-covalent constraints appropriately is thereby detrimental to the success of the rigidity analysis. In this context, rigid regions are those with a well-defined equilibrium structure,[3] whereas biologically important diffusive motion is expected to occur at the flexible regions. At a first sight, rigidity analysis says nothing about the direction and magnitude of existing motions. However, the identification of flexible regions on a bond level gives insights into the location of possible motions or how flexibility characteristics change upon complex formation.[4–6]

Until recently, the approach, implemented into the FIRST software package,[1] has been successfully applied to the protein world, whereas a thorough validation on RNA structures has been missing. Yet, the structural stability of proteins (dominated by hydrophobic interactions) and RNA structures (dominated by hydrogen bonds and base stacking interactions) is determined by different non-covalent forces. Here, we thus aim at developing a network representation of RNA structures that allows for reliably determining flexible and rigid regions within these biomolecules. We then apply rigidity analysis to the large ribosomal subunit to gain insight into the ribosomal exit tunnel's role in co-translational processes.

2 Results

2.1 A Topological Network Representation for Analyzing Flexibility Characteristics of RNA Structures

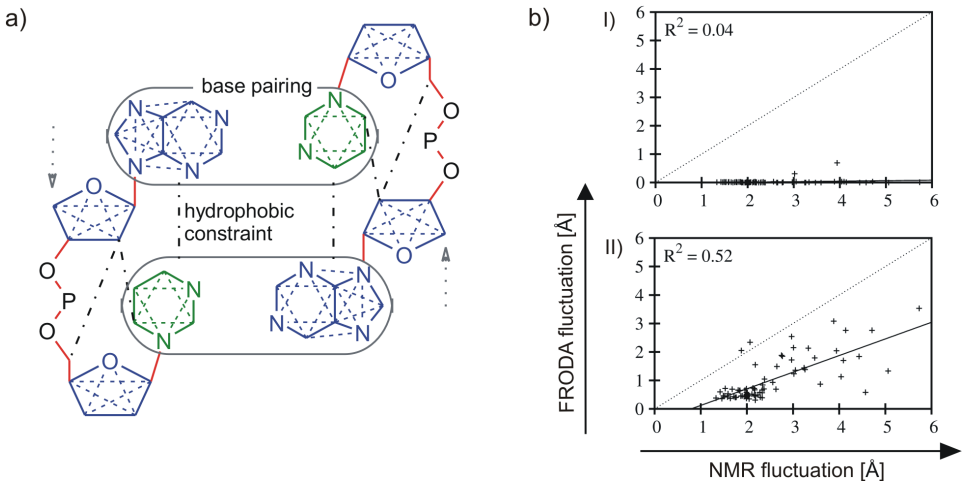


Figure 1. a) Topological network representation of a canonical A-form RNA. Constraints between nearest neighbors are indicated by straight lines, constraints between next nearest neighbors (angle constraints) by dashed lines. For reasons of clarity, angle constraints are only indicated in the sugar and base scaffolds, and hydrogen bonds between bases are omitted. Hydrophobic constraints are indicated by black dashed-dotted lines. Flexible hinges are shown in red, minimally rigid regions in green, and overconstrained regions (which contain redundant constraints) in blue. b) Atomic fluctuations predicted by FRODA simulations[10] vs. conformational variabilities as measured in NMR for RNA structure 1P5O. For the FRODA simulations, a topological network representation according to I) the protein-based parameterization and II) the RNA parameterization was used.

In a topological network representation a constraint is either present or not. To predict reliably the flexibility characteristics of RNA structures, parameters had to be developed for when hydrogen bonds and hydrophobic interactions are included as non-covalent constraints. These parameters were validated based on experimental mobility data of a $tRNA^{ASP}$ structure and all NMR-derived ensembles of RNA structures (with a chain

length $\geq$ 40).[7] We found that restricting the number of base stacking interactions between sequentially adjacent bases to one is crucial and that hydrophobic interactions in general should be considered if the distance of two hydrophobic atoms is smaller than the sum of their van der Waals radii plus a threshold of 0.15 Å. The resulting topological network representation for a canonical A-Form RNA is shown in Figure 1a. Compared to the protein parameterization,[1,4] the new parameters prevent the RNA network representation from being overly rigid (Figure 1b). Further details about the underlying rigidity theory and the validation on RNA and protein structures have been described elsewhere.[1,5,8,9]

2.2 Flexibility Characteristics of the Ribosomal Exit Tunnel Analyzed by Rigidity Theory

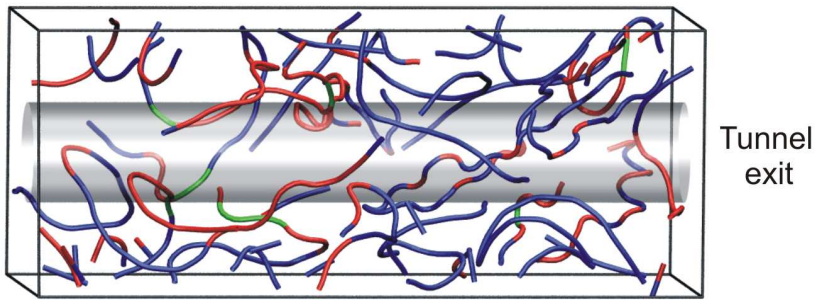


Figure 2. Color-coded representation of the flexibility characteristics of the ribosomal exit tunnel obtained by constraint counting. The coloring of the backbone atoms of the RNA part is according to the flexibility index of the P atoms and according to the C_α-atoms in the protein part. Blue color indicates overconstrained regions and red color flexible regions.

The rigidity analysis can easily be applied to large biomolecules like the ribosomal structure, which consists of more than 10^5 atoms. The computational time amounts to only minutes in this case. The analysis gives new insights in atomic detail into the functional role of the ribosomal exit tunnel, e.g., during protein synthesis. Figure 2 shows a color-coded representation of the flexibility characteristics of the tunnel backbone. Blue color indicates overconstrained regions, green color isostatic (minimally rigid) regions, and flexible regions are colored in red. The approach identifies large parts of the tunnels neighboring regions to be rigid. Remarkably, this holds true for all high resolution structures of the ribosomal exit tunnel of different organism available in the PDB data base.[11–13] Even more striking is the finding of conserved local zones of flexible residues within the tunnel: Clusters of flexible tunnel components are located in the first half of the tunnel and around the tunnel exit. Interestingly, these regions correspond to previously identified folding zones within the tunnel.[14] This striking agreement between tunnel regions with low structural stability and observed folding zones implies that indeed secondary structure may be stabilized entropically there through local conformational adaptability of the ribosomal exit tunnel.

Acknowledgments

We are grateful to financial support by the DFG (SFB 579, "RNA-ligand interactions"), Goethe-University Frankfurt, and Christian-Albrechts-University Kiel. S.F. acknowledges financial support from the Hessian Science Program, the Frankfurt International Graduate School for Science (FIGSS), and the Otto-Stern-School (OSS) for Integrated Doctoral Education in Natural Sciences (Goethe-University Frankfurt).

References

1. D. J. Jacobs, A. J. Rader, L. A. Kuhn, and M. F. Thorpe, *Protein flexibility predictions using graph theory*, Proteins **44**, 150–165, 2001.
2. D. Jacobs and B. Hendrickson, *An Algorithm for Two-Dimensional Rigidity Percolation: The Pebble Game*, J. Comput. Phys. **137**, 346–365, 1997.
3. H. Gohlke and M. F. Thorpe, *A Natural Coarse Graining for simulating Large Biomolecular Motion*, Biophys. J. **91**, 2115–2120, 2006.
4. A. J. Rader, B. M. Hespenheide, L. A. Kuhn, and M. F. Thorpe, *Protein unfolding: rigidity lost*, Proc. Natl. Acad. Sci. U. S. A. **99**, 3540–3545, 2002.
5. H. Gohlke, L. A. Kuhn, and D. A. Case, *Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach*, Proteins **56**, 322–337, 2004.
6. B. M. Hespenheide, A. J. Rader, M. F. Thorpe, and L. A. Kuhn, *Identifying protein folding cores from the evolution of flexible regions during unfolding*, J. Mol. Graph. Mod. **21**, 195–207, 2002.
7. S. Fulle and H. Gohlke, *Analysing the flexibility of RNA structures by constraint counting*, Biophys. J. **94**, 4202–4219, 2008.
8. D. J. Jacobs and M. F. Thorpe, *Generic rigidity percolation: The pebble game*, Phys. Rev. Lett. **75**, 4051–4054, 1995.
9. W. Whiteley, *Counting out to the flexibility of molecules*, Phys. Biol. **2**, 116–126, 2005.
10. S. Wells, S. Menor, B. Hespenheide, and M. F. Thorpe, *Constrained geometric simulation of diffusive motion in proteins*, Phys. Biol. **2**, 127–136, 2005.
11. D. J. Klein, P. B. Moore, and T. A. Steitz, *The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit*, J. Mol. Biol **340**, 141–177, 2004.
12. M. Selmer, C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan, *Structure of the 70S ribosome complexed with mRNA and tRNA*, Science **313**, 1935–1942, 2006.
13. B. Schuwirth, M. Borovinskaya, C. Hau, W. Zhang, A. Vila-Sanjurjo, J. Holton, and J. Cate, *Structures of the Bacterial Ribosome at 3.5 Å Resolution*, Science **310**, 827–834, 2005.
14. J. Lu and C. Deutsch, *Folding zones inside the ribosomal exit tunnel*, Nat. Struct. Mol. Biol **12**, 1123–1129, 2005.

Ab initio Molecular Dynamics of the Zn–Binding Site of the Alzheimer's Amyloid β–Peptide

Sara Furlan[1,2] **and Giovanni La Penna**[1]

[1] Institute for Chemistry of Organo–Metallic Compounds, National Research Council, via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
E-mail: {sfurlan, glapenna}@iccom.cnr.it

[2] Department of Chemistry, IFM, University of Torino, via P. Giuria 7, 50019 Torino, Italy

The aggregation of the peptide amyloid-β (Aβ) into fibrils is considered to be a key event in Alzheimer disease. Zn(II) ion binds the N-terminal segment of Aβ peptide (region 1-16) and influences its aggregation behaviour. Many experimental evidences revealed that Zn-binding involves the three histidine residues (6, 13 and 14). As for the fourth ligand, instead, different candidates have been proposed: Asp 1, Arg 5, Ser 8, Tyr 10 and Glu 11. Here we present the results obtained by *ab initio* molecular dynamics simulations (Car-Parrinello method) at 300 K of Zn(II)-Aβ(1-16) system and implications are discussed.

1 Introduction

In the Alzheimer's disease an accumulation of Aβ peptides occurs and fibrils are formed in the extracellular space of brain tissues. The reason of the accumulation is not known, but the amyloid precursor protein (APP) is probably involved in Cu transport. Other metal ions, like Zn, compete with Cu in APP binding: disfunction of metal ions homeostasis can be at the basis of the pathogenesys. Therefore, the investigation of the Cu and Zn binding sites in APP is important. Both Zn(II) and Cu(II) ions bind the N-terminal segment of Aβ peptide (region 1-16). Among the most recent and mostly debated experimental data, NMR results provided draft structures of Aβ(1-16)-Zn^{+2} complex in aqueous solution at pH=6.5[1] and the Zn binding involves as ligands His 6, His 13, His 14 imidazole groups and Glu 11 carboxylic group in tetrahedral coordination.

The identification of the fourth zinc ligand is extensively debated in literature: Asp 1, Arg 5, Ser 8, Tyr 10 and Val 12 have been also suggested as possible ligands participating in Cu^{+2} or Zn^{+2} chelation by Aβ[2]. The interplay between ligand flexibility and zinc quantum mechanics make the study of zinc coordination in Aβ(1-16) an ideal arena for *ab initio* Car-Parrinello[4] molecular dynamics (CPMD) simulation. The aim of our work is to model the structure of Aβ(1-16)-Zn^{+2}, indicating the fourth ligand.

2 Method

The starting system configurations for CPMD simulations were generated from random walk hybrid Monte Carlo trajectories[5] of a classical model of Aβ(1-16)Zn^{2+} system, employing the Amber force-field modified for possible candidate ligands. We chose four initial configurations with different fourth ligand: Glu 11, Asp 1 (bound to Zn by carboxyl groups), Tyr 10 and Ser 8 (bound by hydroxyl groups). CPMD simulations were made

using a parallel version of the Quantum-ESPRESSO package[3] which incorporates Vanderbilt ultrasoft pseudopotentials and Perdew–Burke–Ernzerhof exchange–correlation. After both electronic and atomic energy minimizations, the system was slowly heated up to 300 K and kept at this temperature for 1.6 ps of molecular dynamics simulation. Representative configurations have been optimized for detailed structural investigations.

3 Results and Discussion

Optimized configurations for the four Zn coordinations show that the three histidines (His 6, His 13, His 14) are bound to Zn. Only Ser 8 is displaced, as fourth ligand, by a backbone carbonyl group. By analysing the time evolution of Zn binding distances measured at 300 K we deduce that: in the single Asp 1 case, one histidine (His 13) is displaced from Zn, and this happens when Asp 7 carboxyl group deprotonates His 14 which then binds more strongly to Zn (distance oscillations in time are less wide). The binding of Zn by carboxyl groups (Asp 1 and Glu 11 cases) is stable. On the contrary, hydroxyl group (Tyr 10 case) binds Zn when assisted by hydrogen bond; as far as this hydrogen bond breaks, hydroxyl group is displaced by nearby peptide carbonyl group, as in the Ser 8 case minimization. Zn is almost always in a near tetrahedral coordination.

Zn-His binding is stable, except for His 13 in the very peculiar case of deprotonation of His 14 by Asp 7. Zn binding by hydroxyl groups of Tyr 10 is possible only when assisted by hydrogen bonds. In solution, the extent of water extrusion from the binding site (favouring intramolecular hydrogen bonds) can stabilize hydroxyl group as a fourth ligand. As far as carboxylic groups can approach Zn, they displace hydroxyl groups and eventual backbone carbonyl groups. This situation is expected to be different when Zn is replaced with Cu: the affinity of Cu for carboxylic and carbonyl groups is lower than that of Zn. Zn can displace Cu by the 1-16 portion of the chain, pushing Cu towards sites involving residues in different Aβ regions[2].

Acknowledgments

This work has been done with the financial support of the FIRB 2003 project RBNE03PX83 of MIUR (Italy). The calculations have been performed on supercomputer IBM cluster JUMP in NIC Supercomputing Centre (Jülich, Germany).

References

1. Zirah S., Kozin S. A., Mazur A. K., Blond A., Cheminant M., Segalas-Milazzo I., Debey P. and Rebuffat S., J. Biol. Chem. **281**, 2151, 2006.
2. Danielsson J., Pierattelli R., Banci L., Graslund A., FEBS Journal **274**, 46, 2007.
3. Baroni S., Dal Corso A., de Gironcoli S., Giannozzi P., Cavazzoni C., Ballabio G., Scandolo S., Chiarotti G., Focher P., Pasquarello A.,Laasonen K., `http://www.pwscf.org/`.
4. Car R. and Parrinello M., Phys. Rev. Lett. **55**, 2471, 1985.
5. La Penna G., Morante S., Perico A., Rossi G. C. J. Chem. Phys. **121**, 10725, 2004.

Do Amyloidogenic Regions Intersect with Folding Nuclei of Native Structure?

Oxana V. Galzitskaya

Institute of Protein Research, Russian Academy of Sciences,
Pushchino, Moscow Region, Russian Federation
E-mail: ogalzit@vega.protres.ru

A crucial event of protein folding is the formation of a folding nucleus, which is a structured part of the protein chain in the transition state. We demonstrate a correlation between locations of residues involved in the folding nuclei and locations of amyloidogenic regions. The average Φ-values are significantly greater inside amyloidogenic regions than outside them. We have found that fibril formation and normal folding involve many of the same key residues, giving an opportunity to outline the folding initiation site in protein chains.

1 Introduction

In spite of the fact that each protein has its own unique, native three-dimensional structure, some cases exist when there is another rather stable structure, called an amyloid fibril. Although native structures vary greatly from protein to protein, the structures of amyloid fibrils obtained from different proteins are fairly uniform. The formation of amyloid fibrils is a case of protein misfolding, in which a protein folds into a cross β-structure instead of folding into its native structure. In addition to proteins that form amyloid fibrils *in vivo* in various "amyloid diseases", there are many other proteins that are not implicated in amyloid diseases but form fibrils *in vitro*[1]. There is no sequence homology common to all such proteins or peptides.

Since polypeptide chains can fold into native structures or misfold into amyloid fibrils, there is a competition between the processes of folding and misfolding. During folding, a protein molecule has to overcome a free-energy barrier. The most unstable structure corresponds to the top of the barrier (i.e., to the transition state of the folding process)[2]. The folding nucleus is a structured part of the protein chain in the transition state. Since the folding nucleus is unstable, it is not easy to investigate it experimentally. A very laborious experimental method, which is called Φ-analysis, has been developed to determine the structure of folding nuclei[2].

The goal of this work is to compare amino acid residues which are crucial for folding and misfolding processes of the same proteins. As the experimental data on both folding nuclei and amyloidogenic regions in the same proteins are scarce, we compared experimentally found residues involved in folding nuclei with predicted residues involved in amyloidogenic regions and *vice versa*. We demonstrate that fibril formation and normal folding involve many of the same key residues. On average, Φ-values for amino acid residues in amyloidogenic regions are significantly greater than Φ-values for amino acid residues in non-amyloidogenic regions. This result allows us to search for some residues involved in the folding nucleus using only amino acid sequences.

2 Results and Discussion

2.1 Intersection of Predicted Residues Involved in Amyloidogenic Regions with Experimentally Found Residues Involved in Folding Nuclei

If amyloid fibril formation is a generic feature of proteins, some common properties of amino acid sequences possessing amyloidogenic propensities should be observed. Therefore, we can hypothesize that amyloidogenic regions often play a crucial role not only in the amyloid fibril formation but also in the process of "normal" folding of proteins into their native structure. We tested whether the experimentally found amino acid residues involved in folding nuclei intersect with theoretically predicted residues involved in amyloidogenic regions. The list of the experimentally found Φ-values as well as the corresponding mutations can be found at Ref. 3. Experimentally found Φ-values (larger 0.5) and predicted amyloidogenic regions can be found at Ref. 4.

We have compared predicted amyloidogenic regions with experimentally found residues involved in folding nuclei for those 20 proteins. The prediction of amyloidogenic regions was made by the previously described method which predicts amyloidogenic regions using only amino acid sequence[5,6]. For each amino acid residue, the method predicts the number of expected contacts and regions within which all residues have a large number of expected contacts are predicted as amyloidogenic ones. As it was demonstrated previously, this method is able to predict amyloidogenic regions[5].

The comparison of the degree of involvement into the folding nucleus (reflected in experimental Φ-values) of residues in the predicted amyloidogenic and non-amyloidogenic regions have demonstrated that there is a reliable difference. In the predicted amyloidogenic regions, the average over Φ-values is 0.41 ± 0.02 while in the predicted non-amyloidogenic regions, the average over Φ-values is 0.33 ± 0.01 (here and below, the shown error is the error of averaging which is calculated as $\frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation of the distribution, and n is the number of points). Student's *t*-test gives the probability of 4×10^{-3}; thus, the above difference is statistically reliable.

Thus, comparison of experimentally known amino acid residues involved in the folding nuclei *vs.* predicted amyloidogenic fragments indicates that nucleation centers for folding and for misfolding often intersect.

2.2 Intersection of Experimentally Determined Amyloidogenic Regions with the Predicted Folding Nuclei

To investigate folding/unfolding behavior of amyloidogenic proteins, we have constructed a database of globular proteins with experimentally revealed amyloidogenic regions. From literature data, we selected those globular proteins in which the position of amyloidogenic regions is known from experimental data.6 The database now includes seven proteins: acylphosphatase, β2-microglobulin, gelsolin, transthyretin, lysozyme, myoglobin, human prion. We tested whether the theoretically found folding nuclei by our method[7] intersect with experimentally found amyloidogenic regions. It appears that 8 of 12 amyloidogenic regions are situated in folding nuclei where Φ-values are large. For several proteins, the regions with the largest Φ-values coincide with the amyloidogenic regions.

For amino acid residues in amyloidogenic regions, the average Φ-value is 0.58 ± 0.02 while amino acid residues in non-amyloidogenic regions have the average Φ-value that

is significantly smaller (0.43 ± 0.01). Thus, in amyloidogenic regions, an average amino acid residue has more than 50% of its contacts formed in the transition state (i.e., in the folding nucleus by our definition). The *p*-value obtained with Student's *t*-test (that is, the probability that the observed difference is accidental) is $2*10^{-11}$ that confirms that the difference between the average Φ-values of amino acid residues in amyloidogenic and in non-amyloidogenic regions is significant.

Thus, we have demonstrated that amyloidogenic regions are often predicted to be part of the folding nuclei in amyloidogenic proteins. Therefore, we can hypothesize that amyloidogenic regions often play a crucial role not only in amyloid fibril formation but also in the process of "normal" folding of amyloidogenic proteins into their native structure, since amyloidogenic regions compose part of the folding nucleus in these proteins.

Acknowledgments

We are grateful to S.O. Garbuzynskiy for assistance in some calculations. This work was supported by the program "Molecular and Cell Biology" of Russian Academy of Sciences, by the Russian Science Support Foundation, by the Russian Foundation for Basic Research, by the INTAS grant (05-1000004-7747), and by Howard Hughes Medical Institute (grant 55005607).

References

1. J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell and C. M. Dobson, *Amyloid fibril formation by an SH3 domain*, Proc. Natl. Acad. Sci. USA **95(8)**, 4224-4228, 1998.
2. J. T. Matouschek, Jr. Kellis, L. Serrano and A. R. Fersht, *Mapping the transition state and pathway of protein folding by protein engineering*, Nature **340(6229)**, 122-126, 1989.
3. `http://phys.protres.ru/resources/phi_values_amyloid.html`.
4. `http://phys.protres.ru/resources/phi_values.htm`.
5. O. V. Galzitskaya, S. O. Garbuzynskiy and M. Yu. Lobanov, *Prediction of amyloidogenic and disordered regions in protein chains*, PLoS Comput. Biol. **2(12)**, 4224-4228, 2006.
6. O. V. Galzitskaya, S. O. Garbuzynskiy and M. Yu. Lobanov,*Is it possible to predict amyloidogenic regions from sequence alone?*, Journal of Bioinformatics and Computational Biology **4(2)**, 373-388, 2006.
7. S. O. Garbuzynskiy, A. V. Finkelstein and O. V. Galzitskaya,*Outlining folding nuclei in globular proteins*, J Mol Biol **336(2)**, 509-525, 2004.

Coarse-Grained Simulations of Protein Adsorption on Solid Surfaces

Beate Griepernau[1], Christian Hanke[1], Ludger Santen[2], and Rainer A. Böckmann[1]

[1] Theoretical and Computational Membrane Biology, Center for Bioinformatics Saar, Saarland University, 66041 Saarbrücken, Germany
E-mail: {b.griepernau, rainer}@bioinformatik.uni-saarland.de

[2] Statistical Physics of Nonequilibrium and Disordered Systems, Theoretical Physics, Saarland University, 66041 Saarbrücken, Germany
E-mail: l.santen@mx.uni-saarland.de

The adsorption of proteins on solids and soft materials plays a vital role in biotechnical and biomedical applications, for example for the biocompatibilty of implant material or in dental health care. Not only the properties of the sorbent surface can be changed, but also the proteins might undergo conformational changes during adsorption. To investigate such processes in molecular detail, but still reaching appropriate time scales (microseconds), coarse-grained molecular dynamics simulations were applied here. As a model system, the adsorption of lysozyme and human serum albumin to a simple, slightly negatively charged, single-layer solid surface were studied at various ion concentrations. Adsorption rates of the two proteins, protein diffusion before and after attachment to the surface, and the orientation of the proteins on the surface were analyzed.

1 Introduction

The modeling of protein adsorption is challenging, because even the single protein dynamics takes place on timescales that are hardly accessible within atomistic simulations. Therefore it is necessary to use coarse-grained models, which describe only partially the internal degrees of freedom of the proteins and the surrounding solution. While the limitations of all-atom simulations are of the order of 10 nm in space and 100 ns in time, they can be extended to the microsecond timescale by the application of a coarse-grained (CG) simulation scheme. In CG simulations, small groups of atoms are treated as single particles, thus reducing the total number of simulated degrees of freedom. Additionally, the larger objects allow for a 10- to 20-fold increase of the time step for integration (from 2fs to 20 – 40fs); also, the absolute dynamics in CG systems is increased by a factor of 4 with respect to real systems[1]. In total, the accessible simulation time of med! ium-sized systems is thereby increased to the submillisecond timescale. Recently, the CG approach has e.g. successfully been applied to study the assembly of transmembrane proteins into lipid bilayers or the self-assembly of high-density lipoproteins.

2 Methods

Coarse-grained simulations on the microsecond time scale were employed to study the aggregation of individual proteins on a negatively charged surface mimicking a mica surface, as well as the lateral protein diffusion. Proteins and water molecules were modeled in the recently developed MARTINI force field[1,2], with additional constraints on the protein to

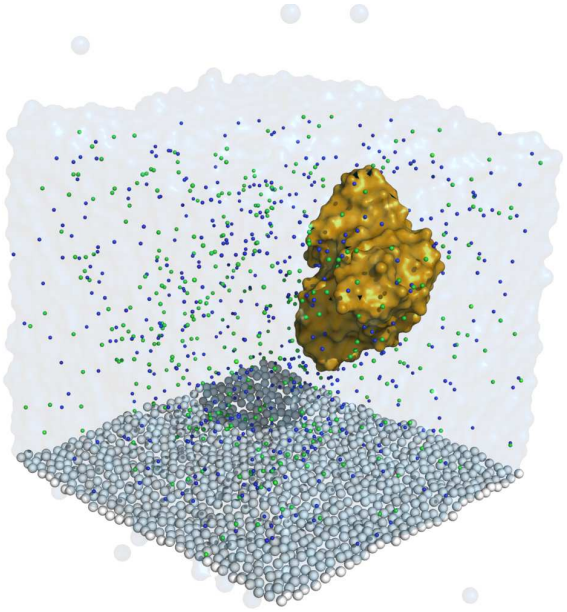


Figure 1. Simulation system for the aggregation study of human serum albumine (HSA, orange) on a solid surface (grey spheres) in explicit ionic solution (0.2 M NaCl, green and blue spheres). The box dimensions are approximately 15 nm in all directions, the total number of cg atoms is approximately 26,000. Aggregation times of HSA to the mica surface range between 1 and 10 μs. Initially, HSA is randomly placed in the water box.

preserve the overall secondary and tertiary structure. Initially, the proteins were randomly placed into the water box (Fig. 1).

3 Results

Typical aggregation times determined from the coarse-grained MD simulations range from 1 μs for lysozyme to about 5 μs for human serum albumine, reflecting the different sizes and therefore different diffusion coefficients of the investigated proteins (164 vs. 585 residues). While these time- and lengthscales (the cg system for human serum albumine corresponds to an all-atom system size of roughly 350,000 atoms) are inaccessible by all-atom simulations, coarse-grained simulations even allow to obtain sufficient statistics on the aggregation process: From in total 27 aggregation simulations of lysozyme on the surface (duration of each run 4 μs) contact patterns of surface amino acids with the solid surface could be determined (see Fig. 2). These clearly show preferred orientations of lysozyme on the surface.

4 Discussion

Protein aggregation on modeled surfaces has been shown to be accessible by a coarse-grained simulation scheme, thereby opening the lane towards atomic-scale studies on the formation of biofilms. Protein diffusion is drastically reduced on the surface, however,

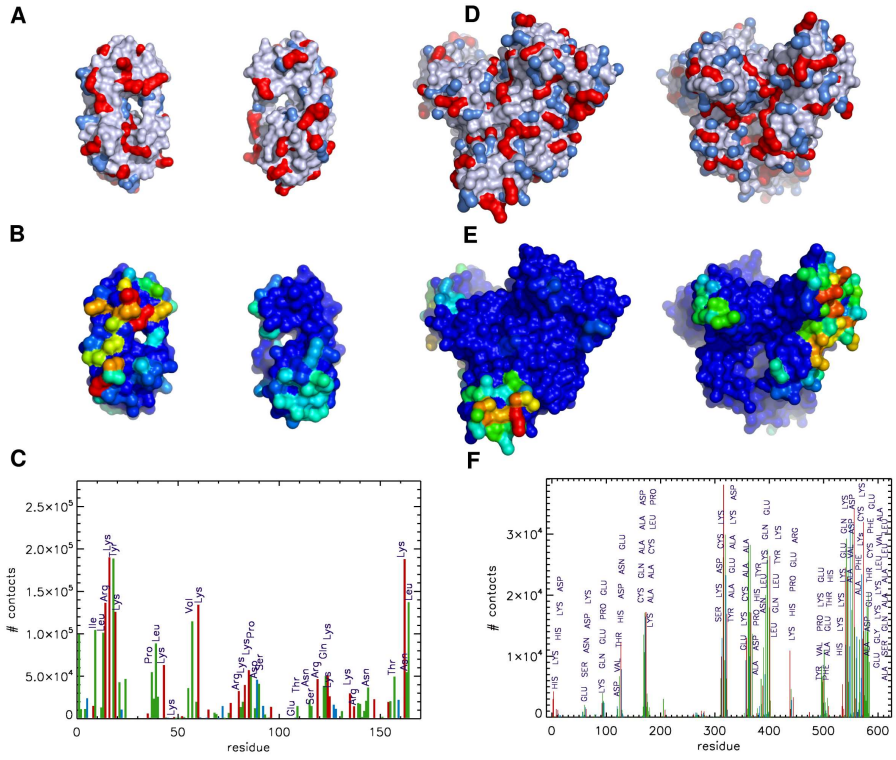


Figure 2. All-atom representation of lysozyme (A) and of HSA (B) with residues colored according to their net charge (red +1, blue -1). Additionally, the coarse-grained proteins (B and E) are shown with color-coded interaction frequencies (C and F) to the negatively charged surface (red high frequency; blue no contacts).

still present. Next steps include partial freeing of the introduced elastic constraints on the protein to allow for large-scale conformational changes and/or unfolding.

Acknowledgments

We thank Karin Jacobs for valuable discussions. Financial support by the Deutsche Forschungsgemeinschaft (Graduate School *Structure Formation and Transport in Complex Systems*, No. 1276/1; BIZ 4/1) is gratefully acknowledged.

References

1. S. J. Marrink, A. H. de Vries and A. E. Mark, *Coarse Grained Model for Semiquantitative Lipid Simulations*, J. Phys. Chem. B **108**, 750–760, 2004.
2. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. Marrink, *The MARTINI coarse grained force field: extension to proteins.*, J. Chem. Theory Comp. **4**, 819–834, 2008.

Towards Understanding the Early Events in the Conformational Transition of Amyloid Beta Peptides

Sikander Hayat and Volkhard Helms

Center for Bioinformatics, Saarland University, 66041 Saarbrücken, Germany
E-mail: {s.hayat, v.helms}@bioinformatik.uni-saarland.de

It is experimentally known that oligomerization of amyloid beta peptides is accompanied by a conformational transition from mainly alpha or random coil to beta sheets. The aim of this study is to analyze and compare the spatial orientation of hydration water near the peptide surface during this conformational transition of amyloid-beta 42 (Ab42) and amyloid-beta 40 (Ab40) peptides. Therefore, molecular dynamics (MD) simulations of 100 ns length with explicit representation of solvent were performed for individual amyloid beta monomers. Analysis was based on the radial distribution function (RDF) of hydration water for individual residues and for respective secondary structure elements. In all cases, initial results suggest that, in accord with the literature, the RDFs reveal the presence of two solvation shells around polar residues. Variations in RDF in the first solvation shell were found to be consistent with the physiochemical properties of the amino acids and were independent of the secondary structure element. However, individual residues that belonged to the secondary structure segments undergoing conformational transitions showed significant redistribution of water density. Further investigations, such as dimer formation and analysis of the orientation of water molecules near peptide surfaces are necessary to clarify the role played by surrounding water in the assembly of such unstructured peptides.

1 Introduction

Recent experimental evidence has implicated the toxicity of soluble oligomers of amyloid beta peptides in Alzheimer's disease[1]. Given the metastable nature of these oligomers, it is hard to obtain experimental data for the early events taking place during the oligomerization of amyloid beta peptides. Computational simulation methods are, hence, needed to provide atomistic details of the early events in amyloid beta oligomerization and there is already a broad literature[2]. As water is known to play a key role in protein folding, structure, dynamics, specific interactions and ligand binding, in this preliminary study, we focus on the spatial organization of water molecules present in solvent surrounding the amyloid beta peptides[3]. Halle recently discussed the technical advancements employed in investigating the influence of protein on surrounding solvent molecules[4], while Helms[5] and Bizzarri[6] reviewed the computational studies aimed at studying protein-water interface and the properties of water near protein surfaces.

The primary aim of this study is to understand the effect of amino acid polarity and peptide secondary structure on the spatial organization of water in its surrounding. To this end two 100 ns MD simulations of Ab40 and Ab42 monomers, respectively, were carried out in explicit solvent. The results obtained from the analysis of radial distribution of water molecules in both the simulations are compared based on the physio-chemical properties of individual residues and the secondary structure composition of residues. Amyloid beta peptides were chosen for this study as they show a well characterized conformational transition and hence can be ideal candidates for studying the effect of secondary structure

elements on surrounding water[7]. The RDFs, MD simulation set up and the starting structure of the peptides is briefly described in section 2. We first discuss the conformational transitions before analyzing the RDFs obtained from both simulations in section 3. The results in section 3 are discussed based on the individual residues, secondary structure elements and with respect to the residues that undergo conformational transition during the course of the simulations. The results presented here are taken from an ongoing investigation and certainly more independent MD simulations need to be conducted to determine the statistical relevance of the results discussed.

2 Methods and Simulation Setup

The initial coordinates for the Alzheimer Ab42 and Ab40 peptides were obtained from the solution structures of the Ab42 peptide in an apolar microenvironment (PDB ID: 1iyt) and of Ab40 in a water-micelle environment (PDB ID: 1ba4), respectively. All standard MD simulations were carried out with the Gromacs package[8] using the GROMOS96 53a5 force field in the NPT ensemble at 300 K and periodic boundary conditions. The linear constraint solver (LINCS) method was used to constrain bond lengths, allowing an integration step of 2 fs. Electrostatic interactions were calculated with the Particle-Mesh Ewald algorithm. **RDFs** describe the ratio between the local density around a reference site r_P and the average density ρ of water molecules in the solution. Here, the terminal atoms of the functional groups of individual residues were considered as reference points r_P and the RDFs were calculated for both the water oxygens $[g_{PO}(r)]$ and water hydrogens $[g_{PH}(r)]$. The RDFs were computed for 1) all amino acids based on their polarity 2) all the amino acids belonging to the same secondary structure element. It is to be noted that in case 2, average RDFs were calculated over certain time intervals to account for the conformational transition taking place during the course of simulation. All residues in the two peptides, irrespective of their solvent exposed surface area, were considered for estimating RDFs. The RDFs were calculated using the g_rdf module of Gromacs[8]. The secondary structure analysis was performed based on DSSP[9].

3 Results and Discussion

We focus here on the effect of the peptide conformational transition on the spatial organization of water surrounding the peptide. First we report the conformational transitions occurring in the peptides during the simulations. In the case of Ab42 peptide, the second alpha helix comprising of residues 28-39 of the Ab42 peptide converted within 5 ns into several beta sheets connected via beta bridges. However, the central hydrophobic region (residues 16-21) remained mostly in 5-helical structure with few local transitions to alpha helical structure during the 100 ns simulation. The remaining residues 1-14 mainly adopted random coil structure and finally settled into a beta sheet conformation at about 40 ns. Also, at about 40 ns, residues 28-33 formed a beta sheet that remained stable until 100 ns. Residues 34-42 formed a beta sheet at about 85 ns that remained stable till the end of the simulation. However, it is to be noted that the beta sheet secondary structure might be favored by the force field applied. In vast contrast to Ab42, no beta sheets formed in the Ab40 peptide and most of the residues remained in random coil conformation for

most of the simulation duration. It is noteworthy to mention that before being replaced by random coil structure, the alpha helix comprising of residues 10-17 remained stable for about 60 ns. The remaining residues 15-36 underwent a conformational transition from alpha helix to random coil within the first few nanoseconds of simulation. Ab40 residues 20-25 and 31-34 remained in random coil structure for most of the duration and adopted a beta sheet structure at about 80 ns. Remarkably, the simulations presented here capture the beta sheet forming tendency of both the peptides. In particular, it is to be noted that Ab42 is considered to be more prone to fibril formation and the emergence of beta sheet structure is suggested to play a key role in its oligomerization[10]. However, the aim of this study is not to highlight this beta sheet formation but rather understand the fluctuations in (solvent) water surrounding the segments that undergoes conformation transition.

The RDFs $[g_{PH}(r)]$ and $[g_{PO}(r)]$ measuring water-hydrogen and water-oxygen density at given distance, respectively, were computed for the terminal atoms of the functional groups present in the two peptides. As expected, the spatial distribution of water was effected by the polarity of the amino acid residue. Two solvent shells were clearly observed for the polar and charged residues, while only one shell was formed around the apolar residues. The water density at the first solvation shell was found to be lower for the apolar residues ($A_{max} \sim 0.7$) as compared to the polar and charged ones ($A_{max} \sim 1.1$). Further, $[g_{PH}(r)]$ and $[g_{PO}(r)]$ values were compared to estimate the orientation of water molecules about the residues. As expected for hydrogen bond donors like ARG and LYS, the first maxima for $[g_{PH}(r)]$ ($r_{max} \sim 2.6$Å) was found to be shifted to a larger distance as compared to $[g_{PO}(r)]$ ($r_{max} \sim 2$Å), signifying the fact that water hydrogens point away from hydrogen bond donor molecules[11].

To investigate the effect of secondary structure transition on water distribution, we identified the residues that had undergone conformational transition and calculated $[g_{PH}(r)]$ and $[g_{PO}(r)]$ functions with reference to the $C\alpha$ atom of the respective residue. In summary, residues 20-25 and 31-34 were identified to have undergone structural transition from random coil to beta sheets in Ab40. In the case of Ab42, residues 1-5 (random coil to beta sheet), 20-24 (alpha helix to random coil) and 34-42 (random coil to beta sheet) were observed to have undergone structural transition. Analysis of RDFs revealed that the effect of structural transition on the water distribution was negligible when the whole segment that had undergone transition was considered. However, the RDFs calculated for individual residues within a given segment seemed to have significant perturbations. As an outlook for this ongoing research, we plan to further investigate these perturbations in water distribution due to secondary structure transition. Moreover, multiple MD simulations with different force fields need to be run to attain statistically relevant results.

Acknowledgments

SH thanks the Graduiertenkolleg 1276/1 for his PhD fellowship.

References

1. Haass, C. and Selkoe, D. J., *Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide*, Nat. Rev. Mol. Cell Biol. **8**, 101–12, 2007.

2. Ma, B. and Nussinov, R., *Simulations as analytical tools to understand protein aggregation and predict amyloid conformation*, Curr. Opin. Chem. Biol. **10**, 445-452, 2006.
3. Massi, F. and Straub, J. E., *Structural and dynamical analysis of the hydration of the Alzheimers-amyloid peptide*, J. Comput. Chem. **24**, 143-153, 2002.
4. Halle, B., *Protein hydration dynamics in solution: a critical survey*, Philos. Trans. R. Soc. London, Ser. B **359**, 1207, 2004.
5. Helms, V., *Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules*, Chem. Phys. Chem. **8**, 23–33, 2007.
6. Bizzarri, A.R. and Cannistraro, S., *Molecular dynamics of water at the protein-solvent interface*, J. Phys. Chem. B **106**, 6617–6633, 2002.
7. Xu, Y., Shen, J., Luo, X., Zhu, W., Chen, K., Ma, J. and Jiang, H., *Conformational transition of amyloid {beta}-peptide*, PNAS **102**, 5403, 2005.
8. Lindahl, E. and Hess, B. and van der Spoel, D., *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, J. Mol. Modeling **7**, 306–317, 2001.
9. Kabsch, W. and Sander, C., *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*, Biopolymers **22**, 2577–2637, 1983.
10. Cheon, M., Chang, I., Mohanty, S., Luheshi, L.M., Dobson, C.M., Vendruscolo, M. and Favrin, G., *Structural Reorganisation and Potential Toxicity of Oligomeric Species Formed during the Assembly of Amyloid Fibrils*, PLoS Comput. Biol. **3**, 1727–1738, 2007.
11. Schröder, C., Rudas, T., Boresch, S. and Steinhauser, O., *Simulation studies of the protein-water interface. I. Properties at the molecular resolution*, J. Chem. Phy. **124**, 234907, 2006.

Protein Interactions with their Environment

Martin Hoefling and Kay E. Gottschalk

Chair for Applied Physics, Ludwig-Maximilians-University Munich, Germany
E-mail: {martin.hoefling, kay.gottschalk}@physik.uni-muenchen.de

During functioning, proteins interact with and influence their environment. The analysis of interactions of proteins with their environment is of crucial importance for an understanding of protein function. Here, we focus on two aspects of protein interactions:

The first topic of interest is the analysis of the protein-protein complexation behavior. In this context, we analyze the complexation and the impact of mutations on this process. As a model system we chose the bacterial ribonuclease Barnase with its natural inhibitor Barstar. Here, our specific interest is the formation of intermediate states along the reaction coordinate as well as the driving complexation force in the system. In agreement with experimental data, we found that the complexation process is mainly driven by electrostatic interactions. This dramatically reduces the conformational space of the approaching complexation partners, resulting in the formation of stable encounter complexes. From these intermediates, the final complex structure is promoted.

The second topic is the modeling of proteins interacting with surfaces in the framework of the ProSurf EU project. In this context our focus is the simulation of protein adsorption on gold surfaces in water. As a first step we evaluated a classical set of parameters derived by ab-initio calculations from our cooperation partners. In obtaining mean force profiles for all 20 amino acids by constrained simulations and comparing them to experimental results, we found reasonable agreement between experimental and computational results. Additionally these simulations allow us to retrieve information first about the amino acid orientation towards the surface during different stages of complexation and second the total free energy difference during adsorption for each amino acid. In our simulations a clear barrier, attributable to the final water layer, could be observed.

1 Introduction

Interactions of proteins with their environment are fundamental for understanding the mechanisms of biological and hybrid systems consisting of biological and inorganic compounds. For transient complexes, electrostatic steering has an important contribution[1] to the association of the proteins. This contribution depends on the distribution of charges across the complexation partners as well as on the properties of the surrounding solvent. Contrary to macroscopic systems, the solvent properties are not homogeneous and isotropic but therefore depend on the surrounding protein surfaces[2].

In biological circumstances, proteins interact not only with their counterpart but also with inorganic surfaces like bone. Compatibility with non-biological surfaces and classification of protein-surface interactions is of increasing importance for nanotechnology and drug design. Yet, a physical understanding for these interactions is currently lacking. The major target of the ProSurf[a] EU project is the development of a toolkit allowing the characterization of protein-surface interactions. In this task we evaluated derived force field parameters in molecular dynamics (MD) simulations.

[a] `http://www.s3.infm.it/prosurf/`

2 Potential of Mean Force

The Potential of Mean Force (PMF) along a chosen reaction coordinate allows the calculation of the free energy difference between two states. The method of choice in our systems is the evaluation of constraint forces[3,4] in simulations with constrained distances along the reaction coordinate. In case of the Barnase-Barstar model system, we chose the distance between the Centers of Mass (COM) as our reaction coordinate. The COM distance of the amino acid from the topmost gold layer plane has been the choice in our gold-amino acid systems. 21 distances (27 in the gold systems) have been sampled with at least $4 \times 5ns$ simulations per distance. The obtained Mean Force profiles are integrated to their potential form.

3 Protein Complexation

3.1 Mutations and Setup

Our target of studies of protein complexation is the well known system consisting of a ribonuclease Barnase and its inhibitor Barstar[5]. To analyze the impact of mutations on the electrostatic steering, we mutated Lys27 and Arg59 on Barnase as well as Asp39 and Glu76 on Barstar to Alanine, as experimentally suggested by Ref. 6 & 7, in one complex of the crystal structure[8]. Simulations at various constrained COM distances were conducted while monitoring the constraint force on the complex constituents as well as the orientation of the water molecules during the simulation. The forces obtained from runs consisting of $5ns$ simulation time at different distances were integrated as described in Sec. 2.

3.2 Results

In our Potentials of Mean Force significant differences between wildtype and mutated complex can be observed. While $\Delta F \approx 60\frac{kJ}{\text{mol}}$ is in reasonable agreement with experimental values[6], the mutated complex shows negligible free energy differences from bound to unbound state compared to the wildtype suggesting a major contribution of electrostatic interaction to the complexation energy difference.

A second observation in our distance constrained simulations was the presence of stable dipole fields at separation distances (additional displacement along COM-COM vector of complexation partners) of $20\mathring{A}$ between complexation partners in analogy to findings in simulations with single peptides[9]. These fields could be observed in wildtype simulations, but not in those with mutants.

4 Protein-Surface Interactions

4.1 Parametrization of 111 Gold Surfaces

The gold surface used in our simulations has been parametrized with the following scheme developed by our ProSurf cooperation partners in Modena: The van-der-Waals interaction of gold atoms is carried by virtual sites in the plane of gold surface atoms. These virtual sites are placed in the geometrical center of each triangle formed by neighboring gold surface atoms. Electrostatic interaction is modeled with dipoles at the position of all gold atoms as described in Ref. 10.

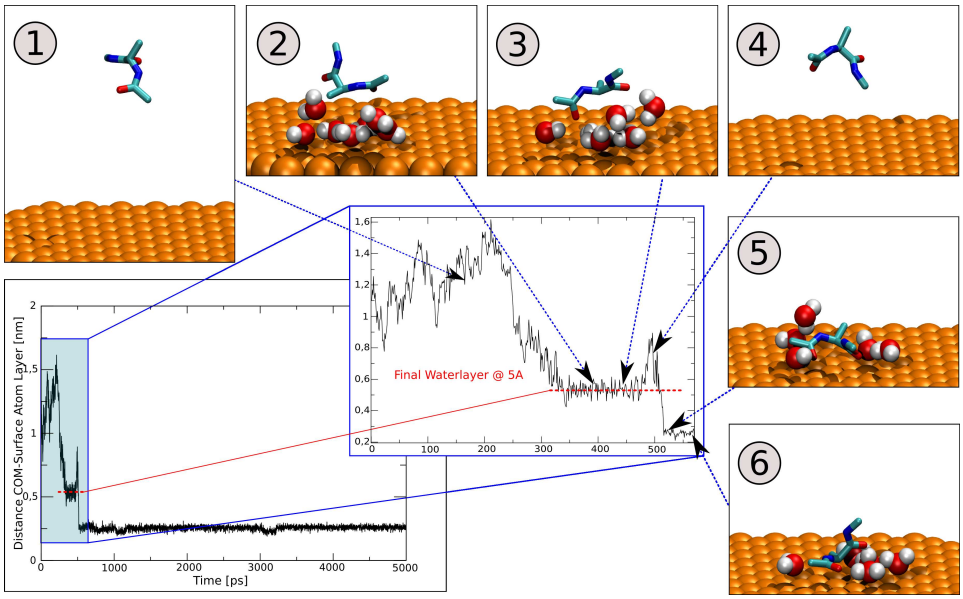


Figure 1. Adsorption process on parametrized gold surface of an amino acid (Alanine) with caps. The free simulation starts at a Center of Mass distance of $1nm$ from surface gold atom layer. A clear barrier at a distance of $5\mathring{A}$ is observable in the time-distance trace and attributable to the final water layer in the corresponding simulation snapshots. Only water within $5\mathring{A}$ of gold atoms and amino acid is shown in the snapshots.

4.2 Free Energy Calculation

In analogy to PMF calculation in Protein Complexes, we calculated Mean Force Profiles for amino acids with capped backbone. Our first results are in agreement with experimentally obtained values from Surface Plasmon Resonance (SPR) measurements and allow us further tuning of gold parametrization in 4.1.

4.3 Dewetting and Adsorption

To adsorb on the gold surface, the final separating water layer, as a barrier, needs to be overcome. Figure 1 shows a typical dewetting process of an amino acid. The barrier at a distance of $5\mathring{A}$ in the time vs. distance trace is clearly visible in frame (2) and (3) as the final water layer. The adsorption process in free simulations starting from $1nm$ distance above the gold surface is very fast for the uncharged amino acids ($\approx 500ps$) while the time until adsorption is significantly longer for charged and polar amino acids. This suggests different barriers during adsorption processes. When adsorbed on the gold surface, we could not observe desorption events in our simulation from any amino acid during $5ns$ simulation time.

5 Concluding Remarks

During complexation, water as the surrounding solvent can mediate a prealignment of the complexation partners, dramatically reducing their conformational space. This is even more surprising since the binding site of Barnase-Barstar is not more hydrophobic than the rest of the protein surface suggesting an increase of electrostatic steering instead of the hydrophobic effect as the major contribution to prealignment. Additionaly, investigations on the adsorption process of amino acids on gold surfaces identify the final waterlayer as primary adsorption barrier.

Acknowledgments

Many helpful discussions with ProSurf team members are gratefully acknowledged. Financial support from the Fonds der Chemischen Industrie, EU Project ProSurf and the Center for NanoScience is acknowledged.

References

1. Alexander Spaar, Christian Dammer, Razif R Gabdoulline, Rebecca C Wade, and Volkhard Helms, *Diffusional encounter of barnase and barstar*, Biophysical Journal, **90**, no. 6, 1913–1924, Mar 2006.
2. Florin Despa, *Biological water: Its vital role in macromolecular structure and function.*, Annals of the New York Academy of Sciences, **1066**, 1–11, Dec 2005.
3. T. Mulders, P. Kruger, W. Swegat, and J. Schlitter, *Free energy as the potential of mean constraint force*, Journal Of Chemical Physics, **104**, no. 12, 4869–4870, Mar. 1996.
4. D. Trzesniak, A. P. E. Kunz, and W. F. van Gunsteren, *A comparison of methods to compute the potential of mean force*, ChemPhysChem, **8**, no. 1, 162–169, Jan. 2007.
5. Robert W. Hartley, *Barnase and barstar: two small proteins to fold and fit together*, Trends in Biochemical Sciences, **14**, no. 11, 450–454, Nov. 1989.
6. C. Frisch, G. Schreiber, C. M. Johnson, and A. R. Fersht, *Thermodynamics of the interaction of barnase and barstar: Changes in free energy versus changes in enthalpy on mutation*, Journal of Molecular Biology, **267**, no. 3, 696–706, Apr. 1997.
7. C. Frisch, A. R. Fersht, and G. Schreiber, *Experimental assignment of the structure of the transition state for the association of barnase and barstar.*, Journal of Molecular Biology, **308**, no. 1, 69–77, Apr 2001.
8. V. Guillet, A. Lapthorn, R. W. Hartley, and Y. Mauguen, *Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar.*, Structure, **1**, no. 3, 165–176, Nov 1993.
9. J. Higo, M. Sasai, H. Shirai, H. Nakamura, and T. Kugimiya, *Large vortex-like structure of dipole field in computer models of liquid water and dipole-bridge between biomolecules*, Proceedings of The National Academy of Sciences of The United States of America, **98**, no. 11, 5961–5964, May 2001.
10. F. Iori and S. Corni, *Including image charge effects in the molecular dynamics simulations of molecules on metal surfaces.*, J Comput Chem, Mar 2008.

Identification of Differential Protein Expression in Response to the Application of BioRegulators that Enhance Plant Productivity and Quality

Ruth Horn, Denise Zimmermann, and Stefan Schillberg

Plant Biotechnology, Fraunhofer IME, Aachen, Germany
E-mail: ruth.horn@molbiotech.rwth-aachen.de

The enhancement of plant growth and productivity by the use of chemicals is a common practice in agriculture but rather little is known about the effect these BioRegulators have on plants on a molecular level. The presented work is part of an ERA-NET project with the aim to identify genes, proteins and metabolites, differentially expressed under conditions of stress while treated with certain compounds. Arabidopsis thaliana plants are grown on soil and in hydroponic cultures and are subjected to salt, cold and drought stress. The progress of stress is monitored phenotypically by leaf shape and chlorophyll fluorescence. The proteome profile of the plants is analysed via 2D electrophoresis and mass spectrometry. For the identification of differentially regulated proteins in response to the BioRegulator treatment under stress conditions, the DIGE (difference-in-gel-electrophoresis) system will be used. The biomarkers identified universally for the different stress conditions will be used to establish a cell based assay for the screening of potential new BioRegulator compounds.

1 Plant Growth

Arabidopsis plants are grown on soil for the study of cold stress and in hydroponic cultures for the study of salt and drought stress. In both setups plants are grown for three weeks under short day conditions (9 hrs light, 15 hrs dark) and one week under long day conditions (16 hrs light, 8 hrs dark) before they are subjected to the respective stress conditions. For cold stress conditions the plants are shifted to a day/night temperature regime of 4 °C. Control plants stay at the day/night temperature regime of 21°C/18°C. Salt stress is produced by supplementing the hydroponic solutions with 125 mM NaCl and drought stress by supplementing the solution with 20 % PEG, thereby lowering the water potential.

2 Stress Monitoring

Stress monitors had to be found in order to detect significant changes of plant behaviour under stress conditions while treated with BioRegulator compounds. For cold stress the decreasing quantum yield of photosynthesis (probed via chlorophyll fluorescence of PSII) can be used as a reliable indicator for stress. Healthy plants exhibit a ratio of 0.8 whereas a decrease below 0.7 indicates disturbance and damage of the photosynthetic apparatus by the stress. The onset of salt stress is monitored in changes of leaf shape as in comparison to control plants, the leaves of stressed plants grow rather in width than in length. The ratio between leaf length and width was measured at certain time points during a salt stress period of 144 hrs. At the end the ratios were 2.22 ± 0.14 and 2.78 ± 0.24 for stressed and control plants respectively. Drought stress leads to a reduced overall growth and a

darker green colouring compared to normal conditions but these parameters are so far only documented photographically. Using the accumulation of anthocyanin, a purple coloured secondary metabolite as a stress monitor is currently tested.

3 Compound Treatment

BioRegulator compounds are sprayed onto the plants 24 hrs before they are shifted to the respective stress conditions. Among others, compounds from the groups of azoles and neonicotinoids are used in the study.

4 Proteomic Analysis

2D electrophoresis is used for the analysis of the whole leaf proteome. Proteins extracted from Arabidopsis leaves are separated according to their isoelectric point (pI) in the first dimension by isoelectric focusing (IEF) and according to their molecular weight in the second dimension by SDS polyacrylamid gel electrophoresis (PAGE). For the identification of the biomarkers 2D electrophoresis using fluorescent protein tags (DIGE, difference in gel electrophoresis) is used for the identification of the biomarkers. With this method, two samples from different conditions to be compared are labelled with fluorescent dyes that are spectroscopically distinguishable (CyDyesTM GE Healthcare) and run on the same 2D gel. An internal standard consisting of a pool of all samples present in the experiment is run on every gel in order to facilitate inter-gel matching and statistic evaluation (three samples per gel, CyDyes 2, 3 and 5). Spot intensities from the different images are obtained and protein difference ratios are determined using the DeCyder software (GE Healthcare).

First the difference in protein composition in response to cold stress conditions untreated with BioRegulators was investigated by DIGE. Arabidopsis plants were subjected to cold stress for a period of 144 hrs and leaf samples (in triplicates) were taken every 24 hrs. In total 8 gels were run. After 144 hrs of cold stress a total of 112 protein spots were found to be differently regulated. Of these, 64 spots increased and 48 spots decreased in abundance. In total ~ 1300 spots were detected on the 2D map of the whole leaf proteome. Protein identification will be done by LC-MS.

In the next steps plants showing a significant effect when treated with BioRegulator compounds under stress conditions (monitored by the parameters introduced in part 2) will be investigated by DIGE in order to indentify the biomarkers necessary for the establishment of the screening assay.

Acknowledgments

This work is supported by the BMBF (project number FKZ 0313989B).

Hamiltonian Replica Exchange Molecular Dynamics Using Soft-Core Interactions to Enhance Conformational Sampling

Jozef Hritz and Chris Oostenbrink

Leiden/Amsterdam Center for Drug Research (LACDR), Division of Molecular Toxicology,
Vrije Universiteit, Amsterdam NL-1081 HV, The Netherlands
E-mail: {hritz, c.oostenbrink}@few.vu.nl

We present a novel Hamiltonian replica exchange molecular dynamics (H-REMD) scheme that uses soft-core interactions between those parts of the system that contribute most to high energy barriers. The advantage of this approach over other REMD schemes is the possibility to use a relatively small number of replicas with locally larger differences between the individual Hamiltonians. Because soft-core potentials are almost the same as regular ones at longer distances, most of the interactions between atoms of perturbed parts will only be slightly changed. Rather, the strong repulsion between atoms that are close in space, which is in most cases resulting in high energy barriers, is weakened within higher replicas of our proposed scheme.

The presented approach leads to a significant enhancement of conformational sampling both for the smaller molecules GTP and 8-Br-GTP in explicit water and for residue Phe483 within the catalytic site of CYP2D6 in complex with its substrate MAMC.

1 Introduction

Replica exchange molecular dynamics (REMD) has shown a tremendous impact in the field of biomolecular simulation. While temperature REMD (T-REMD) is mostly used to enhance conformational sampling of larger systems in implicit solvent, Hamiltonian REMD (H-REMD) is also suitable for simulations in explicit solvent. However, it is not always trivial to find a perturbation of the Hamiltonian which leads to enhanced conformatioinal sampling.

Here we present a H-REMD scheme using soft-core interactions[1] which is particularly suitable for the enhanced sampling of selected flexible parts of systems in which the energy barrier between different conformations is high due to strong non-bonded repulsions. The barriers are lowered by weakening the interactions using soft-core interactions given by eq. (1):[2]

$$V_{LJ} = \left(\frac{C12}{\alpha C12/C6\lambda^2 + r^6} - C6 \right) \frac{1}{\alpha C12/C6\lambda^2 + r^6} \tag{1}$$

All simulations were performed using GROMOS05[3] and the GROMOS 53A6 force field.The REMD efficiency was significantly increased by allowing multiple replicas to run at the highest softness level. Values of other softness levels were optimized by mimicked REMD to maximize the number of global conformational transitions.[4]

2 REMD of GTP and 8-Br-GTP

We have tested the new protocol on the GTP and 8-Br-GTP molecules in explicit solvent, which are known to have high energy barriers between the anti and syn conformation of the

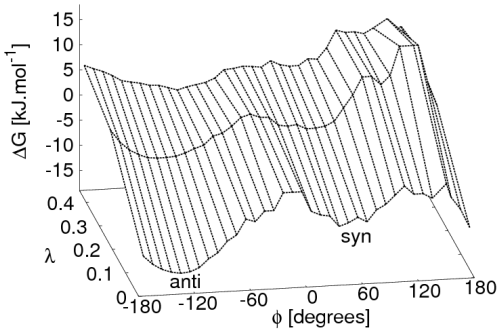


Figure 1. Potential of mean force along the glycosidic bond of GTP, as function of the softness parameter λ.

base with respect to the sugar moiety. During two 25 ns MD simulations of both systems the transition from the more stable to the less stable (but still experimentally observed) conformation is not seen at all. Also T-REMD over 50 replicas for 1 ns did not show any transition at room temperature. On the other hand, more than 20 of such transitions are observed in our new H-REMD scheme using 6 replicas (at 3 different Hamiltonians) during 6.8 ns per replica for GTP and 12 replicas (at 6 different Hamiltonians) during 7.7 ns per replica for 8-Br-GTP. The soft-core interactions were applied for the non-bonded base-sugar interactions in GTP and 8-Br-GTP.

The calculated population of GTP in the anti conformation was 95.6% $\pm$ 0.5% and 6.0% $\pm$ 1.8 % for 8-Br-GTP.[1] These values are in very good agreement with results coming from thermodynamic integration using a hidden dihedral angle restraint around the glycosidic bond. The observed inverse character for GTP and 8-Br-GTP is also in agreement with NMR estimates for anti/syn populations. Dihedral angle distributions around the glycosidic bond were also used to generate the potential of mean force (Fig. 1). From this figure it can be seen that the energy barrier as well as the free energy difference between the anti and syn conformation is decreasing with increasing softness. This leads to an increased number of conformational transitions at these levels of softness.

3 REMD of Phe483 within CYP2D6

Cytochromes P450 (CYPs) are heme-containing enzymes that can be found in virtually all organisms. CYP2D6 is one of the most crucial isoforms involved in the drug metabolism of humans. Mutagenesis experiments confirm importance of Phe483 in substrate binding. Its conformation is however quite unclear as it is positioned in a rather flexible loop region inside the catalytic site of CYP2D6. Several computational studies indicate that multiple Phe483 sidechain conformations are occuring. 10 ns of CYP2D6 MD simulations in complex with several ligands revealed only very few transitions for Phe483 between conformations corresponding to $\chi^1 = 70^\circ$ and $\chi^1 = 170^\circ$. Visual inspection indicates

that the corresponding energy barrier is mostly due to the repulsion between the Phe483 and Leu224 side-chains together with the dihedral angle term around χ^1. Therefore enhanced conformational sampling by H-REMD using soft-core interactions was performed for Phe483 within CYP2D6 in complex with the MAMC substrate in which the force constant for the χ^1 dihedral angle term was additionally decreased towards zero with increasing softness. When softness was applied only for the interactions between the sidechains of Phe483 and Leu224 no enhanced sampling was observed. Rather, the sidechains moved closer in space and a high energy barrier remained. On the other hand, when softness was applied for all interactions involving the Phe483 sidechain and the rest of the protein or MAMC, the barrier was significantly lowered. We observed 9 global conformational transitions within 1 ns of H-REMD using 8 replicas. These preliminary (not fully converged) simulations reveal that the conformation of Phe483 with $\chi^1 = 70^\circ$, as it is observed in the (apo) crystal structure is present for only 15% of the time when MAMC is bound in the active site. Interestingly, χ^1 had value $\sim 170^\circ$ in the homology model of CYP2D6 in complex with the codein, which was constructed in our group before crystal structure was released.

4 Concluding Remarks

We have developed a H-REMD scheme using soft-core interactions[1] and implemented it into the GROMOS05 package. Its high conformational sampling efficiency is shown for GTP and 8-Br-GTP as well as for Phe483 within the CYP2D6/MAMC complex. The high efficiency is obtained thanks to the fact that only those parts of the Hamiltonian are perturbed which contribute most to high energy barriers. Another efficiency gain was obtained by using a degenerate highest softness level and the optimal H-REMD settings obtained from optimization by REMD mimicking.[4]

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research, VENI Grant No. 700.55.401.

References

1. J. Hritz, and C. Oostenbrink, *Hamiltonian Replica Exchange Molecular Dynamics Using Soft-Core Interactions*, J. Chem. Phys. **128**, 144121, 2008.
2. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, *Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations*, Chem. Phys. Lett. **222**, 529-539, 1994.
3. M. Christen, P. H. Hunenberger, D. Bakowies, R. Baron, R. Burgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren, *The GROMOS software for biomolecular simulation: GROMOS05*, J. Comput. Chem. **26**, 1719-1751, 2005.
4. J. Hritz, and C. Oostenbrink, *Optimization of Replica Exchange Molecular Dynamics by Fast Mimicking*, J. Chem. Phys. **127**, 204104, 2007.

TollML: A Database of Toll-Like Receptor Structural Motifs

Ferdinand Jamitzky[1,2], Jing Gong[2], Tiandi Wei[2], Wolfgang M. Heckl[2,3], and Shaila C. Rössle[2]

[1] Leibniz Supercomputing Centre, Garching, Germany
E-mail: jamitzky@lrz.de

[2] Department für Umwelt- und Geowissenschaften, LMU München, Germany

[3] Deutsches Museum, München, Germany

During recent years Toll-like receptors (TLRs) have spearheaded a tremendous research interest and the amount of sequenced relevant proteins grows exponentially. A critical step towards the successful TLR structure modeling is to generate the leucine-rich repeats (LRR) motif aided sequence alignment between the target sequence and the templates. However, because of the irregularity of LRR motifs in TLRs, most TLRs have no LRR annotaions in current databases, and in those TLRs with LRR partitions, the indicated repeat number and the boundaries of LRRs are quite different among databases. In order to provide a useful platform for structure prediction and analysis of these sequences, we developed TollML, an XML based database specialized for TLR structural motifs. Its original TLR sequences were extracted from NCBI's protein database. The LRR motifs as well as transmembrane and TIR motifs for all known TLR sequences are identified and annotated manually and can then be used for the prediction of protein structures via alignment and threading. The resulting database has been used for the structure prediction of TLR7, 8 and 9.

1 Introduction

Toll-like receptors (TLRs) play a crucial role in innate immunity[1]. To date, 13 TLRs have been identified in mammalian, and equivalent forms of many of these have been found in other vertebrate species. Under a structural view, the TLRs consists of 3 parts: the TIR domain inside the membrane, the transmembrane region and the ectodomain (ECD) formed by 18 to 25 leucine-rich repeats (LRRs). It is just the ECD that is directly involved in recognition of a variety of pathogens (ligands). Although for most of these TLRs ligand recognition specificity, downstream adapter molecules and signaling pathway have now been established, we still do not know much about their structural interaction with ligands. In 2005 the crystal structure of TLR3 ectodomain (ECD) was resolved and recently the crystal structures of TLR1/2 and 4 in respective complexes with agonist and antagonist ligands were shown. All these explained how the LRR based platform is adapted to the recognition of ligands. However, with high throughput genome sequencing projects the amount of sequenced TLR proteins continues to grow exponentially. It is clear that the discrepancy between the rate at which novel protein sequences are discovered and the rate at which detailed structural information will be obtained from X-ray diffraction or nuclear magnetic resonance spectroscopy (NMR) will continue for the foreseeable future. For this reason, there is a pressing need for theoretical methods to predict protein structures from their sequence. The understanding of their structural interactions can help us design vaccines, understand autoimmune diseases, and define the correlates of immune

protection. The TLR7, 8 and 9 constitute the TLR7 family which is one of the six major vertebrate TLR families. They are all located in the endosome and recognize nucleic acids. Our objective is to construct models for TLR7, 8 and 9 ECD based on the structure known TLRs and other LRR containing proteins.

2 Database Construction

TollML entries were originally extracted from NCBI proteins database and PDB[2] using the search keys: toll* and tlr*, where the star (*) stands for any suffix. The data were then filtered semi-automatically to exclude TLR related molecules such as adaptors, protein kinases and transcription factors. The metabolic pathways information was then extracted from the KEGG-database for each TLR entry[3]. Aside from the extracted information, TollML contains additional annotations, which must be manually accomplished. These annotations include LRR partitions, ligand information and structural information achieved through other projects or extracted from published articles. The indicated number of LRRs and their boundaries in individual TLRs are quite different among databases or researchers. This difference reflects the irregularity of LRR motifs in TLRs. We standardized the LRR definition and partitioned each TLR ECD into LRRs manually. The generated XML-datafile was then stored in an XML-database[4].

3 Contents of Database Entries

The current version (1.2) contains 2232 entries[5]. 1529 entries are of mammalian and the rest are of non-mammalian. All entries are divided into 24 groups, from TLR1 to TLR23 and others. A special tag named TollML label records which group an entry belongs to as a quick search index. The entry distribution over different TLR families is illustrated in Tab. 1. Each entry in TollML provides information of one TLR protein.

TLR	1	2	3	4	5	6	7	8	9	10-14	other	total
Mammalia	55	92	77	921	47	46	71	55	83	81	2	1529
Non-mam	267	132	22	16	22	1	13	8	12	54	159	703
Total	322	224	99	937	69	47	84	63	95	135	161	2232

Table 1. The entry distribution of TLR families for mammalian/non-mammalian groups.

4 Construction of a Conformational LRR Database

Leucine-rich repeats are an array of 20 to 30 amino acid long protein segments. Every segment is rich in the hydrophobic amino acid leucine. They play an important role in protein-protein interactions, such as signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, ice nucleation, apoptosis and

innate immune response. The first crystal structure of LRR containing protein, a ribonuclease inhibitor, was determined in 1993. It is a horseshoe- shaped structure containing 15 LRRs with a parallel beta-sheet lining the inner circumference and alpha helices flanking the outer circumference.

LRRs are present in over 6000 proteins from viruses to eukaryotes. For more than 80 of them the structure is known. In order to create a convenient workbench to carry out the homology modeling and to manage the structure known LRRs flexibly and efficiently, we decided to construct a conformational LRR database.

All leucine-rich repeats can be divided into a highly conserved segment (HCS) and a variable segment (VS). The HCS consists of an 11 or 12 residue stretch with consensus sequence LxxLxLxxN(Cx)xL, in which L stands for Leu, Ile, Val or Phe, N stands for Asn, Thr, Ser or Cys and x is any amino acid. A short beta-sheet begins always at the third position. 4 L residues at position 1, 4, 6, and 11 participate in the hydrophobic core. The side chains of asparagines (N) at position 9 form hydrogen bonds between neighbor LRRs in the loop structure. The VS of LRRs is quite different in length and consensus sequence. It can contain a variety of secondary structures.

5 Protein Comparative Modeling

At present there are over 80 LRR containing proteins whose crystal structures are available in RCSB Protein Data Bank (PDB), including the ECDs of TLR1/2, 3 and 4. They provide useful resources for homology modeling of TLR7, 8 and 9.

Comparative modeling, also called homology search, exploits the fact that evolutionarily related proteins with similar sequences, have similar structures. The process of building a comparative model is conceptually straightforward. First, Blast searches for the sequence to be modeled (the target) against a database of known protein structures is performed to find a most similar sequence (the parent). The similarity is usually greater than 35%. Second, an alignment is generated between the target and the parent. This sequence alignment is used to construct an initial model (sometimes referred to as a framework or template) by copying over some main chain and side chain coordinates from the parent structure based on the equivalent residue in the sequence alignment. At last, the model is improved by energy minimization and molecular dynamics.

Acknowledgments

This work was supported by Graduiertenkolleg 1202 of Deutsche Forschungsgemeinschaft (DFG).

References

1. Takeda K, Akira S, *Toll-like receptors in innate immunity*, Int. Immunol. **17 1**, 1-14, 2005.

2. Helen M. Berman, et al. *The Protein Data Bank*,Nucleic Acids Res. **28**, 235-242, 2000.
3. Gong J, Wei T, Jamitzky F, Heckl W M, Roessle S C, *TollML a User-Editable Database for Toll-like Receptors and Ligands*, Proc. Suppl. 2nd IAPR Itermational Workshop on Pattern Recognition in Bioinformatics (PRIB 2007) , Singapore, 2007.
4. `http://exist-db.org/`
5. `http://zeus.krist.geo.uni-muenchen.de/~tollml`.

The Glucan, Water Dikinase - A Kinetic Model to Understand the Initial Step in Starch Mobilization in Plant Leaves

Önder Kartal[1,2] **and Oliver Ebenhöh**[1,2]

[1] Institut für Biochemie und Biologie, Universität Potsdam, 14469 Potsdam, Germany

[2] Max-Planck-Institut für Molekulare Pflanzenphysiologie,
Wissenschaftspark Golm, 14476 Potsdam
E-mail: {*kartal, ebenhoeh*}*@mpimp-golm.mpg.de*

A novel kinetic model for the first enzymatic step committed to the starch degradation pathway in plant leaves is presented and analyzed. It is based on current knowledge and hypotheses about the action and role of glucan, water dikinase (GWD) an essential enzyme for normal starch breakdown. The model is formulated in terms of ordinary differential equations and includes (a) the autoactivation of GWD, (b) the adsorption/desorption to the starch surface, and (c) the interfacial catalysis of the phase transition from crystalline to amorphous sites. It is this transition only, which renders the insoluble starch granule susceptible to hydrolytic attack by exo- and isoamylases, ultimately releasing soluble glucans into the stroma. We identify the efficiency of phase transition as being important for the final state of the system and the adsorption as a potentially feasible regulatory site to control the rate of phase transition and thus starch breakdown.

1 Introduction

Transitory starch in leaves is an important carbon source for sink organs of plants during the night phase. It is composed of amylose and amylopectin with a ratio of approximately 1/5 to 1/9 depending on botanical origin. This composition renders starch rather hydrophobic and insoluble. The granule is degraded during the night by amylolytic enzymes hydrolizing α-1,4-bonds (mainly by β- amylases) and α-1,6-bonds (isoamylases) to release glucans into the stroma.[3] Prior to hydrolytic attack, however, the bonds have to be made accessible. The transition of inaccessible crystalline sites in the interface to accessible amorphous sites is catalyzed by at least one dikinase, the glucan, water dikinase GWD. This enzyme is supposed to unwind glucan double helices at crystalline sites and phosphorylate glucosyl residues to increase the hydrophilicity of starch, thereby preventing spontaneous helix formation after dissociation. This is consistent with an increase of the phosphate content in starch granules at the beginning of the night and a higher activity of amylases on starch in vitro if GWD is provided.[1,2]

2 Motivation

With the model we were able to simulate possible behavior for different conditions and test for the dependence of the phase transition rate on presumably important parameters like the adsorption equilibrium constant. From hereon, we can generate testable quantitative hypotheses to aid and interpret in vitro assays of GWD on starch granules or crystallised maltodextrins, as well as to improve and especially extend our model to include glucan release.

3 Model Formulation

A simple model of the hypothetical initial enzymatic step in starch breakdown is presented. It comprises four reactions describing the catalytic cycle of GWD. The corresponding system of ordinary differential equations is derived.

3.1 Reaction Mechanism

We assume the following simplified reaction scheme:

$$\begin{array}{ll} E + ATP \rightleftharpoons C_1 \rightarrow AMP + C_2 & \text{(autophosphorylation)} \\ C_2 + S_C \rightleftharpoons \Theta & \text{(adsorption/desorption)} \\ \Theta \rightarrow P_i + \alpha S_A + (1-\alpha) S_C + E & \text{(phase transition)} \end{array}$$

At first, ATP is bound reversibly by GWD (E) to build a complex C_1. In the second reaction the β-phosphate of ATP is transferred to a His-residue in the protein yielding the complex C_2 and AMP. After this autoactivation, GWD is adsorbed in the form of C_2 at crystalline sites S_C on the surface. The variable Θ denotes crystalline sites occupied by the enzyme, which can either dissociate again or catalyze the (partial) phase transition to amorphous starch (S_A) in the fourth reaction . This last step lumps together the transfer of one phosphate group to a glucosyl residue, releasing the other phosphate P_i to the bulk phase and desorption of the enzyme. The parameter $\alpha \in (0, 1)$ is introduced in order to cope with the fact that crystalline sites are possibly not transformed into amorphous sites stoichiometrically. Depending on the length of the glucans forming the double helices more than one phosphate may be necessary to get an amorphous site. If for example $\alpha = 0.5$, on average two ATP are consumed for a complete phase transition at a given site.

We assume that the surface of the starch granule is always either crystalline, amorphous or occupied by enzymes. It has to be emphasized that the rates depending on moieties associated with the surface are assumed to be proportional to the corresponding surface area normalized to the maximally available surface area S. This is a usual assumption, which is for example also used in the kinetic derivation of the Langmuir adsorption isotherm.[4]

3.2 Differential Equations

The time evolution of the particle numbers in mole per time follows

$$\frac{d}{dt}\vec{n} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & (1-\alpha) \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 1 \end{pmatrix} \vec{v}$$

where $n_1 = [ATP] \cdot V$, $n_2 = [E] \cdot V$, $n_3 = [C_1] \cdot V$, $n_4 = [C_2] \cdot V$, $n_5 = [AMP] \cdot V$, $n_6 = S_C \cdot A_{max}$, $n_7 = \Theta \cdot A_{max}$, $n_8 = S_A \cdot A_{max}$, and $n_9 = [P_i] \cdot V$. The squared

brackets denote concentrations whereas V is the volume of the bulk phase. A_{max} is a parameter characteristic for the adsorbed particle, denoting the maximal GWD adsorption in mol/m^2 if the substrate is in excess.

The rate vector $\vec{v} = (v_1, \ldots, v_4)^T$ was derived assuming mass action kinetics. Simplification of the system description is possible by exploiting two conservation relations, $S \cdot A_{max} = n_6 + n_7 + n_8$, and $N = n_1 + n_3 + n_5$. Eliminating two variables, dividing all equations by V and nondimensionalizing time yield the system

$$
\begin{aligned}
\dot{c}_1 &= \tilde{k}_{-1}(c_1^0 - c_1 - c_4) - \tilde{k}_1 c_1 c_2 \\
\dot{c}_2 &= \tilde{k}_{-1}(c_1^0 - c_1 - c_4) + \tilde{k}_4 \cdot W_{max}(1 - s_c - s_a) - \tilde{k}_1 c_1 c_2 \\
\dot{c}_3 &= \tilde{k}_2(c_1^0 - c_1 - c_4) - W_{max}(K_{ad} \cdot c_3 \cdot s_c - (1 - s_c - s_a)) \\
\dot{c}_4 &= \tilde{k}_2(c_1^0 - c_1 - c_4) \\
\dot{c}_5 &= W_{max} \cdot \tilde{k}_4(1 - s_c - s_a) \\
\dot{s}_c &= (1 - \alpha) \cdot (1 - s_c - s_a) - K_{ad} \cdot c_3 \cdot s_c \\
\dot{s}_a &= \alpha \cdot \tilde{k}_4(1 - s_c - s_a)
\end{aligned}
$$

The variables c_1 to c_5 are the concentrations of ATP, GWD, the complex C_2, AMP, and P_i, respectively. The dimensionless surface ratios are $s_c = S_C/S$ and $s_a = S_A/S$ from which $\theta = \Theta/S$ may be calculated by $\theta = 1 - s_c - s_a$. In the simulations, no AMP is provided initially, therefore c_1^0 denotes the initial ATP concentration. The parameter $W_{max} = \tilde{A}_{max} \cdot \frac{m}{V}$, with m being the mass of starch provided is related to both the space occupied by a molecule of GWD on the starch surface and the shape of the granule. The parameter $K_{ad} = k_3/k_{-3}$ in M^{-1} is the equilibrium adsorption constant of GWD to crystalline sites.

4 Results

4.1 Progress Curves

Figure 1 shows a typical time evolution of the system if enough ATP is provided initially. The crystalline zone is completely transformed into amorphous zone. If the ATP level is too low or the phase transition is inefficient (low α) ATP becomes depleted before transition of the surface is complete (data not shown). This would severely disrupt further breakdown of the starch granule in vivo.

4.2 Control Over the Maximum Rate of Phase Transition

To infer which of the initial reaction steps and associated parameters presumably have a strong impact on the breakdown rate of starch we calculated the influence of some of these parameters on the maximum rate of substrate provision for hydrolysis, namely the maximum phase transition rate $\max v_4$. The results depicted in Figure 2 indicate that regulating the adsorption/desorption constant K_{ad} may be a good means to effectively alter v_4. Also the autophosphorylation step of GWD can be a good regulation site, at least if it is sufficiently slow (data not shown).

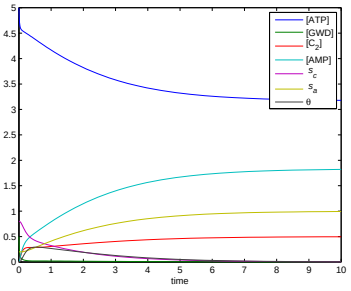


Figure 1. Progress curves of a system with ATP excess. The phase transition is complete and this steady state is stable for many parameter values (data not shown).

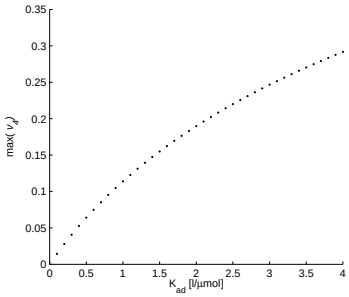


Figure 2. Maximal phase transition rate depending on adsorption equilibrium constant. The positive slope reflects a strong control over the transition rate.

Acknowledgments

We are indebted to M. Steup from the University of Potsdam for intensively discussing with us the peculiarities of starch biochemistry and focusing our attention to critical points concerning the degradation pathway. We also acknowledge financial support from the BMBF via the FORSYS initiative and the facilities at the Max-Planck-Institute of Molecular Plant Physiology.

References

1. C. Edner, J. Li, T. Albrecht, S. Mahlow, M. Hejazi, H. Hussain, F. Kaplan, C. Guy, S. M. Smith, M. Steup, G. Ritte, *Glucan, Water Dikinase Activity Stimulates Breakdown of Starch Granules by Plastidial β-Amylases*, Plant Physiol **145**, 17–28, 2007.
2. G. Ritte, A. Scharf, N. Eckermann, S. Haebel, M. Steup, *Phosphorylation of Transitory Starch is increased during degradation*, Plant Physiol **135**, 2068-2077, 2004.
3. A. M. Smith, S. C. Zeeman, S. M. Smith, *Starch Degradation*, Annu Rev Plant Biol **56**, 73-97, 2005.
4. Y.-H. P. C. Zhang, L. R. Lynd, *Toward an Aggregated Understanding of Enzymatic Hydrolysis of Cellulose: Noncomplexed Cellulase Systems*, Biotechnol Bioeng **88**, 797-824, 2004.

Engrailed Homeodomain Folds Overnight by 100 Processors

Konstantin V. Klenin[1,2] **and Wolfgang Wenzel**[1,2]

[1] Institut für Nanotechnologie, Forschungszentrum Karlsruhe,
Postfach 3640, 76021 Karlsruhe, Germany

[2] DFG-Centrum für Funktionelle Nanostrukturen, Universität Karlsruhe,
Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany
E-mail: {*klenin, wenzel*}*@int.fzk.de*

It has been shown that the native structure of a small protein can be efficiently found as the global minimum of a certain all-atom forcefield. In the present study, we use this approach to simulate folding of Engrailed homeodomain (PDB code 1ENH) containing the helix-turn-helix motif. The search procedure is based on the Monte Carlo simulated annealing combined with the evolutionary algorithm. We propose a new way of increasing the efficiency of this method.

1 Introduction

The free-energy approach has delivered promising results for protein folding and structure prediction in recent years. Following Anfinsen's hypothesis[1], the native state is postulated to be the global minimum of a all-atom free-energy function. This minimum can be found by optimization methods[2] involving the Monte Carlo simulated annealing. This approach was successfully used to fold α-, β- and mixed proteins of small length[3–5].

This procedure requires an accurate, transferable energy function, such as the Protein Force Field (PFF02) developed in our group[6,7]:

$$E = E_{\mathrm{LJ}} + E_{\mathrm{ES}} + E_{\mathrm{SASA}} + E_{\mathrm{HB}} + E_{\mathrm{R}}.$$

It contains the following terms: (1) E_{LJ}, the standard Lennard-Jones potential, (2) E_{ES}, the Coulomb energy of electrotatic interaction with efficient dielectric constants, (3) E_{SASA}, a term proportional to the solvent accessible surface aria, (4) E_{HB}, a term for the hydrogen bonding, (5) E_{R}, a term stabilizing the β-regions in the Ramachandran plots. All the atoms are explicitly represented (only the apolar group CH_n is considered as a single atom). The bond angles and the bond lengths are fixed. The degrees of freedom considered are the backbone (ψ, ϕ) and the sidechain (χ_i) dihedral angles. The solvent is taken into account implicitly.

The Monte Carlo procedure involves two kinds of moves: (1) unbiased, consisting of a ramdom change of the dihedral angles, and (2) biased, which set the dihedral angles (within a single residue) to predefined values from a certain library.

The most promising optimization method is the evolutionary algorithm[8], where a fixed size population of conformations evolves simultaneously. One cycle of the algorithm consists of the following steps. An individual conformation is taken randomly from the population and is subjected to the Monte Carlo simulated annealing starting from a random temperature. The obtained structure is then added to the population. If the similar (in terms of RMSD) conformations are already present, than the only one among them with

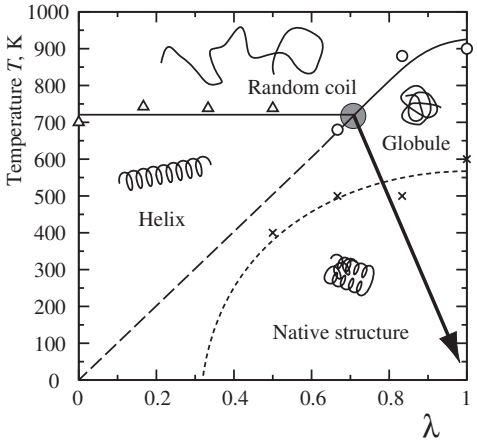


Figure 1. The phase diagram for the Engrailed homeodomain. The optimal annealing path is shown by the arrow.

the lowest energy is kept in the population and the others are removed. Otherwise, the highest energy structure within the entire population is removed (provided the population size has exceeded the certain limit). After many such cycles, the lowest energy structure approaches the global minimum.

In the present work, we improve this procedure by optimizing the conditions of the simulated annealing. The efficiency of our approach is demonstrated by the example of the Engrailed homeodomain (PDB code 1ENH), a small three-helix protein containing the helix-turn-helix motive that is involved in the DNA binding[9].

2 Methods

It has been observed that during the simulated annealing, as the temperature lowers down, the collapse of the chain occurs first, and only then the secondary structure forms. But in the collapsed conformation the Monte Carlo moves are essentially less efficient. To address this problem, we used the modified energy

$$E = \lambda E_{\mathrm{SASA}} + E_{\mathrm{HB}} + ...,$$

where λ is an arbitrary parameter in the range $0 \leq \lambda \leq 1$. Consider the (artificial) phase diagram in the coordinates λ and the temperature T as shown in Fig. 1. (By the "phase" we imply the most probable state of the system.) The "triple" point is characterized by the most extensive fluctuations of the structure in terms of its size and energy. The main idea of this work is to use it as the starting point for the annealing process. (The unrealistic high temperature is due to the stabilizing effect of the biased moves.) First, the system is kept at the fixed λ and T for a period of approximately one relaxation time τ and then, for the same time period, it is moved along the arrow in Fig. 1.

The triangle and circle points in Fig. 1 were obtained without knowledge of the native structure as illustrated in Fig. 2.

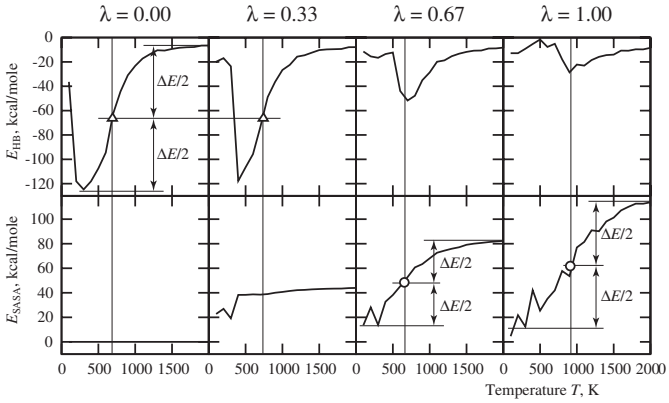


Figure 2. The mean energies E_{HB} and E_{SASA} (for the quasi-stable states reached by the Monte Carlo process starting from an extended conformation) as functions of the temperature T for various λ. The boundary of the helical phase corresponds to the middle value of E_{HB} at $\lambda = 0$. The boundary of the globular phase corresponds to the middle value of E_{SASA} at each particular λ.

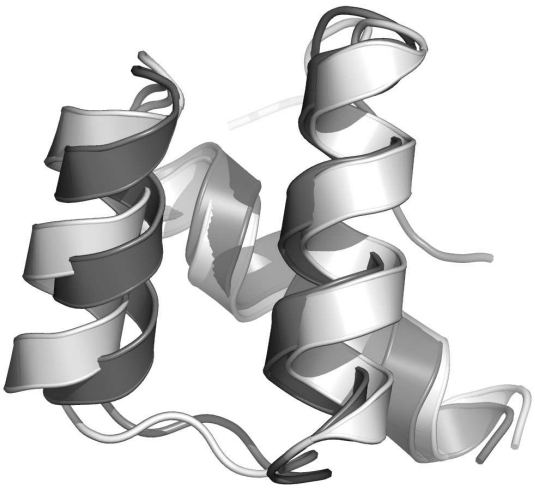


Figure 3. Comparison between the simulated (dark gray) and experimental (light gray) structures. The backbone RMSD is 2.8 Å.

3 Results and Conclusions

The comparison between the simulated and experimental structures is shown in Fig. 3. When the "triple" point of the phase diagram is known, the native structure of the Engrailed homeodomain can be found in 2 months CPU time. Using the optimized annealing path increases the efficiency of the evolutionary algorithm by an order of magnitude.

Acknowledgments

We thank the BMBF, the Fond der Chemischen Industrie, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-1, WE 1863/10-2, WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. We are thankful to the Barcelona Supercomputer Center and to the Korea Institute of Science and Technology for computational resources.

References

1. C. B. Anfinsen, *Principles that govern the Folding of Protein Chains*, Science **181**, 223, 1973.
2. A. Schug, A. Verma, W. Wenzel, and G. Schoen, *Biomolecular structure prediction with stochastic optimization methods,* Adv. Eng. Materials **7**, 1005, 2005.
3. A. Verma, S. Murthy, K. H. Lee, E. Starikov, and W. Wenzel, *De novo all atom folding of helical proteins*, NIC Series **34**, 45, 2006.
4. W. Wenzel, *Predictive folding of a β hairpin in an all-atom free-energy model*, Europhys. Letters **76**, 156, 2006.
5. S. M. Gopal and W. Wenzel, *De Novo Folding of the DNA-Binding ATF-2 Zinc Finger Motif in an All-Atom Free-Energy Forcefield*, Angew. Chemie Int. **45**, 7726, 2006.
6. T. Herges and W. Wenzel, *An All-Atom Force Field for Tertiary Structure Prediction of Helical Proteins*, Biophysical Journal **87**, 3100, 2004.
7. A. Verma and W. Wenzel, *Towards a universal Free Energy Forcefield for All atom Protein Folding*, (Submitted) 2008.
8. A. Verma, S. M. Gopal, J.S. Ooh, K.H. Lee, and W. Wenzel, *All atom de-novo protein folding with a scalable evolutionary algorithm*, J. Comput. Chem. **28**, 2552, 2007.
9. N.D. Clarke, C.R. Kissinger, J. Desjarlais, G.L. Gilliland, C.O. Pabo, *Structural studies of the engrailed homeodomain*, Protein Sci. **3**, 1779, 1994.

Single Molecule FRET Study of the Conformational Energy Landscape of a Transfer RNA

Andrei Yu. Kobitski[1], **Martin Hengesbach**[2], **Mark Helm**[2], **and G. Ulrich Nienhaus**[1]

[1] Institute of Biophysics, University of Ulm, 89069 Ulm, Germany
E-mail: andrei.kobitski@uni-ulm.de, uli@uiuc.edu

[2] Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, 69120 Heidelberg, Germany
E-mail: martin.hengesbach@web.de, mark.helm@urz.uni-heidelberg.de

Local, post-transcriptional modifications can have pronounced effects on the global energy landscape of RNA molecules. We have studied the equilibrium between discrete conformational states in human mitochondrial lysine transfer RNA (mt $tRNA^{Lys}$) by using single-molecule Förster (or fluorescence) resonance energy transfer (smFRET) spectroscopy. From histograms of the FRET efficiency, an unfolded structure (***U***), a non-functional, extended hairpin structure (***E***) and the functional cloverleaf form (***C***) of human mt $tRNA^{Lys}$ can be distinguished. The equilibria between the ***U***, ***E***, and ***C*** states were characterized as a function of Mg^{2+} concentration for two RNA constructs that only differ by a single methyl group modification of a nucleotide base. A thermodynamic model was developed which is based on the separation of conformational changes and binding of divalent cations. Based on this model, the impact of a single methyl group modification on the energy landscape of $tRNA^{Lys}$ was assessed.

1 Introduction

Proteins and ribonucleic acids (RNAs) are linear polymers that can fold into compact three-dimensional structures, in which they are able to perform specific roles in biological processes within living cells or organisms. Finding the correctly folded structure is an extraordinarily complex process that has yet to be solved by *in-silico* modeling approaches, although significant progress has been made over the years[1]. The key problem for computation is the vast conformational space of even moderately sized biopolymers[2]. The conformational energy landscape has provided a conceptual framework by which to describe both protein folding and function[3]. Protein folding is visualized by a transition of the molecular ensemble on the energy landscape via many parallel trajectories and local minima en route to the folded-state ensemble. Folding of proteins and RNA is overall governed by the same principles; yet there are differences arising from the nature of the interactions introduced by the monomeric units. A distinct difference, however, is the hierarchical nature of RNA folding that originates from the pronounced base pairing leading to the formation of relatively stable secondary structures[4].

Frequently, post-transcriptional chemical modifications of ribonucleotides are observed which only slightly change the energy landscape to selectively stabilize the native conformation[5]. Here we present an exception, human mitochondrial (mt) lysine transfer RNA ($tRNA^{Lys}$), in which a single methylation on adenosine 9 (m^1A9) in its structural core was seen to cause a marked shift of the thermodynamic equilibrium toward the functional form of the RNA molecule[5,6]. We have studied this biologically important modification by using single-molecule Förster (or fluorescence) resonance energy transfer

(smFRET), a technique that allows conformational changes of biomolecules to be visualized under equilibrium conditions in real time[7]. It relies on the ability of a fluorescent dye (donor) to transfer its energy non-radiatively to another fluorescent dye (acceptor) that typically absorbs at longer wavelengths. By attaching such a FRET pair of dyes specifically to the structure of interest, the strong distance dependence (R^{-6}) of the effect enables distance changes down to 1 Å to be measured. Such quantitative experimental data are expected to be most useful in the development of both theories and computational modelling approaches.

2 FRET Measurements

Two suggested secondary structures of tRNA^{Lys}, the non-functional extended hairpin (***E***) and the functional cloverleaf-based L-shape (***C***) conformations, are presented in Fig. 1a. To observe the conformational changes between ***E*** and ***C***, induced by the methylation of adenosine 9 (A9), two FRET constructs with the unmodified (Kwt) and the modified (Km^{1}A) sequences were prepared[6]. The presence of multiple conformations in the FRET-labeled RNA constructs was investigated by smFRET experiments on freely diffusing and surface-immobilized molecules using a confocal microscope. In these experiments, the efficiency of FRET, $E = \frac{I_A}{I_A+\gamma I_D}$, is calculated ratiometrically from the donor (I_D) and (I_A) acceptor fluorescence photon counts, and the parameter γ accounts for differences in the donor and acceptor quantum yields and the detection efficiencies. RNA is a polyanion, carrying one unit of charge on each nucleotide, and the Coulomb repulsion must be screened by counterions, which may bind in specific locations or just form a diffuse cloud, to stabilize the functionally competent states. Variation of the counterion concentration is a useful experimental control parameter to shift equilibria between different structures in the energy landscape. We have measured histograms of the FRET efficiency at 16 different Mg^{2+} concentrations, which can be described by superpositions of three FRET efficiency distributions peaking at low, intermediate and high FRET values. Based on the proposed structural model for tRNA^{Lys} and the crystal structures of tRNA^{Phe}, we have assigned these subpopulations to the ***U*** (for "unfolded"), ***E*** (for "extended hairpin"), and ***C*** (for "cloverleaf-based L-shape") states. Switching events between the ***E*** and ***C*** states, with

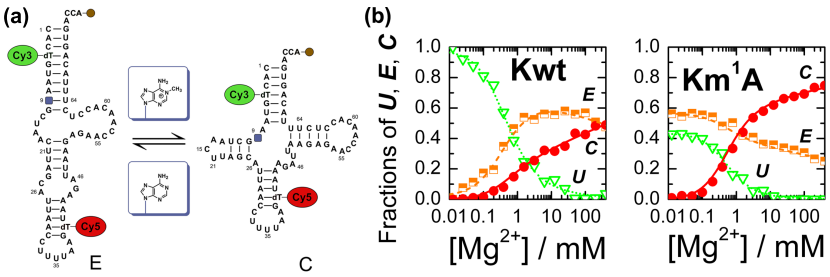


Figure 1. Conformational changes of human mt tRNA^{Lys} affected by methylation of adenosine 9 (square). (a) Secondary structures of the proposed extended hairpin (***E***) and the cloverleaf-based L-shape (***C***) conformations. (b) Mg^{2+} dependence of the fractional populations in the unfolded (***U***), ***E***, and ***C*** states of Kwt and Km^{1}A tRNA^{Lys} constructs. Results from fitting the data with the thermodynamic model are given as lines.

rare brief sojourns to ***U*** states, were observed in FRET trajectories of individual molecules immobilized on a glass surface. At 10-mM Mg^{2+} concentration, the kinetics of $tRNA^{Lys}$ was described with a simple model involving two conformations, ***E*** and ***C***, interconverting on the 100-ms timescale[6]. To quantify the change of the ***U***, ***E***, and ***C*** conformations with Mg^{2+} concentration, we performed a global fit of the 16 FRET histograms for each of the constructs; in Fig. 1b, the fractional populations of the three states of Kwt and Km^1A are plotted. For both constructs, a pronounced drop of the ***U*** state population is observed at $\sim$0.5 mM Mg^{2+}. It is accompanied by an increase of the ***C*** state population with increasing Mg^{2+} concentration. The ***E*** state is much more populated at low concentrations in Kwt than in Km^1A. At high Mg^{2+} concentration ($\sim$100 mM), the ***E*** and ***C*** state populations decrease and increase with Mg^{2+}, respectively.

3 Thermodynamic Model

Ion-induced stabilization of RNA can be modeled by decomposing the reaction into RNA folding and ion binding, as proposed by Misra and Draper[8]. Based on this approach, we have developed the six-state thermodynamic model depicted in Fig. 2a. In this model, there are free energy differences between the Mg-free U_0, E_0, and C_0 conformations, whereas the strength of Mg^{2+} ion binding to the U_{Mg}, E_{Mg}, and C_{Mg} conformations governs the equilibrium between the Mg-free and Mg-bound populations of the corresponding states. Mathematical details of the thermodynamic model can be found in a recent publication[6]. The lines in Fig. 1b represent the fit results of the fractional populations governed by the equilibrium coefficients. The free energies of the Mg-free and Mg-bound (at 1 M) states are depicted in Fig. 2b. The compaction of the conformations going from U_0 to E_0 and from E_0 to C_0 increases the free energies of the Kwt construct. For the Km^1A construct, the methylated A9 introduces a positive charge, which influences base pairing[9] and stabilizes the E_0 state significantly ($\sim$10 kJ/mol). However, the stabilizing effect is drastically reduced for the E_{Mg} of Km^1A, possibly because of the competition between Mg^{2+} binding to the ***E*** state and favorable hydrogen bonding of m^1A9 in the base pair m^1A9-U64.

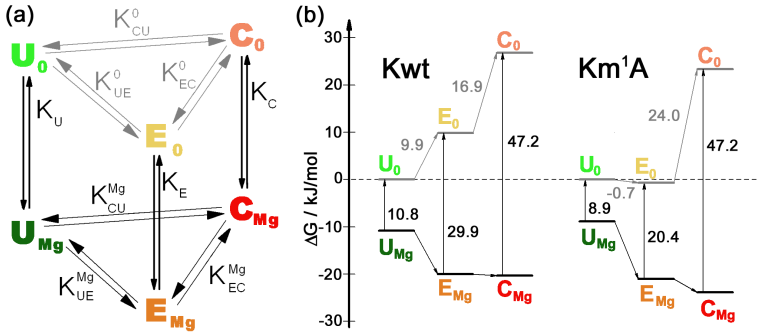


Figure 2. Mg^{2+}-induced $tRNA^{Lys}$ folding reaction. (a) Thermodynamic scheme describing the equilibria between the Mg-free and Mg-bound forms of the ***U***, ***E***, and ***C*** states. (b) Free energy diagram of the populations U_0, E_0, and C_0 of Mg-free, and U_{Mg}, E_{Mg}, and C_{Mg} (at 1-M Mg^{2+}) of Mg-bound Kwt and Km^1A $tRNA^{Lys}$.

In the cloverleaf conformation, A9 is likely exposed to the solvent and, thus, the stabilizing effect is smaller ($\sim$3 kJ/mol). Nevertheless, binding of Mg^{2+} ions remains practically unaffected, making this functional conformation thermodynamically preferable at physiologically relevant ion concentration.

4 Concluding Remarks

Here we have shown that FRET experiments performed at the single molecule level provide details about structural and dynamic aspects of RNA molecules. Equilibria and rate coefficients of conformational transitions can be studied by selectively changing environmental parameters, as was shown in here by variation of the counterion concentration. Quantitative information extracted from such measurement will be most interesting for comparison with computational approaches aimed at simulating these transitions in the complex RNA energy landscape.

References

1. W. F. van Gunsteren and J. Dolinc, *Biomolecular simulation: historical picture and future perspectives*, Biochem. Soc. Trans. **36**, 11-15, 2008.
2. A. V. Finkelstein and O. B. Ptitsyn, *Protein Physics*, Academic Press, San Diego, 2002.
3. J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, *Theory of protein folding: the energy landscape perspective*, Annu. Rev. Phys. Chem. **48**, 545-600, 1997; K. A. Dill and H. S. Chan, *From Levinthal to pathways to funnels*, Nature Struct. Biol. **4**, 10-19, 1997; G. U. Nienhaus and R. D. Young, *Protein dynamics*, in: Encyclopedia of Applied Physics, G. L. Trigg (Ed.), VCH Publishers, New York, Vol. **15**, pp. 163-184, 1996.
4. P. Brion and E. Westhof, *Hierarchy and dynamics of RNA folding*, Annu. Rev. Biophys. Biomol. Struct. **26**, 113-137, 1997.
5. M. Helm, *Post-transcriptional nucleotide modification and alternative folding of RNA*, Nucleic Acids Res. **34**, 721-733, 2006.
6. F. Voigts-Hoffmann, M. Hengesbach, A. Yu. Kobitski, A. van Aerschot, P. Herdewijn, G. U. Nienhaus, and M. Helm, *A methyl group controls conformational equilibrium in human tRNA^{Lys}*, J. Amer. Chem. Soc. **129**, 13382-13383, 2007; A. Y. Kobitski, M. Hengesbach, M. Helm, and G. U. Nienhaus, *Sculpting an RNA conformational energy landscape by a methyl group modification - a single-molecule FRET study*, Angew. Chem. Int. Ed. **47**, 4326–4330, 2008.
7. G. Bokinsky and X. Zhuang, *Single-molecule RNA folding*, Acc. Chem. Res. **38**, 566-573, 2005; G. U. Nienhaus, *Exploring protein structure and dynamics under denaturing conditions by single-molecule FRET analysis*, Macromol. Biosci. **6**, 907-922, 2006.
8. V. K. Misra and D. E. Draper, *The linkage between magnesium binding and RNA folding*, J. Mol. Biol. **317**, 507-521, 2002.
9. P. E. Agris, H. Sierzputowska-Gracz, and C. Smith, *Transfer RNA contains sites of localized positive charge: carbon NMR studies of [13C]methyl-enriched Escherichia coli and yeast tRNA^{Phe}*, Biochemistry **25**, 5126-5131, 1986.

Impact of Induced Fit on Ligand Scoring and a Strategy of Identifying a Minimal Set of Flexible Residues

Daria Kokh[1] **and Wolfgang Wenzel**[2]

[1] FB.C - Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, 42097 Wuppertal, Germany
E-mail: kokh@uni-wuppertal.de

[2] Institute für Nanotechnologie, Forschungszentrum Karlsruhe GmbH, Postfach 3640, 76021 Karlsruhe, Germany
E-mail: wenzel@int.fzk.de

Although conformational changes in receptor upon ligand binding are a very common phenomenon, incorporating protein flexibility in a docking procedure encounters significant computational problems. A possible solution is inclusion side-chain flexibility for only limited number of residues in the binding pocket, which can improve notably docking accuracy without considerable increase of computational costs. However, investigation of this approach is often limited to specifically chosen receptors and mostly focused on the impact of receptor flexibility on docking accuracy, whereas ligand scoring, the real weakness of the present-day docking methodology, is treated only peripherically. In the present study we investigate enrichment rates of rigid-, soft-, and flexible- ("induced-fit") -receptor models using 12 diverse proteins with receptor-specific ligand libraries containing up to 13000 molecules, comprising known ligands and decoys with similar physical properties but distinct topology. We also present and test a straightforward protocol for the choice of the flexible residues, which is based on the ability of the receptor structure to accommodate the set of known ligands. This strategy is an unbiased approach to identify the most important residues likely to be relevant for induced fit effects, which allowed us to improve EF_1 values by $\sim$35% on average with respect to rigid-docking.

1 Method

FlexScreen[1] is an all-atom docking approach based on the stochastic tunneling method[2] for the energy minimization and a simple, first principle based atomistic scoring function that contains a sum of the Van-der-Waals, electrostatic Coulomb, and angular dependent hydrogen bond. The Van-der-Waals parameters have been taken from OPLSAA[3], the partial charges of the receptors have been computed with MOE[4], the hydrogen bond parameters have been taken from AutoDock[5]. The method enables continues rotation up to 15 side chain bonds of the receptor in the energy optimization procedure.

Scoring performance of the Flex-Screen approach has been benchmarked by using 12 target proteins of the DUD database[6] with relatively small binding cavities that are completely buried from solvent. For each target the database includes a set of annotated ligands (up to 350) and a set of decoys containing about 36 molecules for each ligand that resemble the particular ligand in physical properties, but differ topologically, so that they unlikely to be binders. The following receptor have been analyzed: Androgen receptor, Cyclooxynase 1, Cyclooxynase 2, Estrogen receptor agonist, Glycogen phosphorylase beta , Glutacorticoid receptor, Mineralcorticoid receptor, Purine nucleoside phosphorylase, Progesterone receptor, Retinoic X receptor alpha, S-adenosyl-homocysteine hydrolase, and Thymidine kinase.

Docking and screening performance has been evaluated by computing the enrichment of annotated ligands among the top-scoring molecules of the receptor-specific database

$$EF_{\alpha} = \frac{(concentration\ of\ known\ ligands\ found\ in\ top - ranking\ subset)}{(concentration\ of\ known\ ligands\ in\ database)} \quad (1)$$

where EF_{100} equals the fraction of annotated ligands that bind to the receptor in the docking calculations (binding energy less than zero) and, therefore, shows the efficiency of ligand docking, whereas EF_1 indicates screening efficiency with the maximum possible value of 37.

In order to treat receptor flexibility for all systems at an unbiased level, some uniform scheme for choosing the flexible residues has to be implemented. In the present study potentially flexible residues have been identified by requiring that majority of the known ligands must have a negative binding energy to the receptor. We have, therefore, analyzed a list of ligands that are unbound in rigid-receptor docking for a specific target and found a set of residues that most often cause energy clashes with these ligands (vdW energy above 20kJ/Mol). These residues have been ranked according to the number of clashes and form a list of flexible residues. Finally, several top-ranking residues from this list have been treated as flexible.

2 Results

We have found significant limitations of rigid receptor models that for some targets fail to bind even 50% of the known ligands to the apo-structure of the protein. The enrichment rate does not correlate with docking performance and is good only for 4 receptors :ER-agonist, COX2, MR, PNP (see Fig.1). For the other receptors enrichment rates remain poor, in high correlation with a previous study of the same database[6].

We have investigated a soft- and flexible-receptor approximations, by shifting or optimizing (soft- and flexible-docking, respectively) of side-chains from the list of flexible residues to adapt binding pocket for annotated ligands.

As can be expected, both models are effective in finding binding poses for ligands that do not dock in the rigid-receptor calculations, whereby the fraction of annotated ligands that bind to a receptor (EF_{100}) increases monotonically with the number of shifted/flexible residues as illustrated in Fig.2. The variation of enrichment rate, however, is not monotonic and reaches its maximum at about 3-8 flexible/shifted residues for major receptors. Since the energy correction accounting for receptor reconstruction is omitted in soft-receptor model, it is not surprising that this method is not so successful with regards to the enrichment performance: EF_1 do not even reach the values obtained in rigid docking for most receptors (Fig.1, left panel). Unlike soft-receptor, flexible-receptor model increases enrichment rates in comparison with the rigid receptor model for 11 from 12 targets (Fig.1, right panel). The scoring performance of flexible-receptor docking is good ($EF_1 > 20$) for 8 of 12 targets (in comparison to 4 in the case of rigid-docking) and medium ($10 < EF_1 < 20$) in the remaining 4 cases. In contrast to rigid-receptor docking, where for 4 targets screening results are unsatisfactory , we now find $EF_1 > 10$ for all targets.

These results show that accommodation of ligand-induced protein reconstruction by rotating of receptor side chains that are most often involved in steric clashes between a protein and known ligands can notably improve performance of visual screening. The

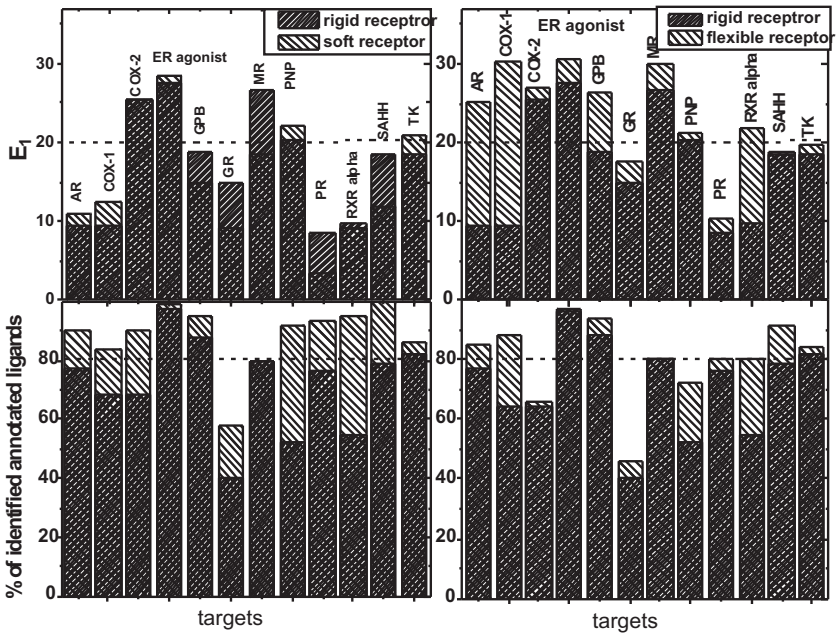


Figure 1. Comparison of enrichment factors of rigid- and soft- (flexible)-receptor models for 12 receptors. The numbers of flexible residues are: 8 (AR), 9 (COX1), 3 (COX2), 4 (ER-agonist), 6 (GPB), 6 (GR), 6 (MR), 3 (PNP), 7 (PR), 11 (RXR-alpha), 2 (SAHH), and 3 (TK). In soft-docking selected residues are shifted by 0.25nm.

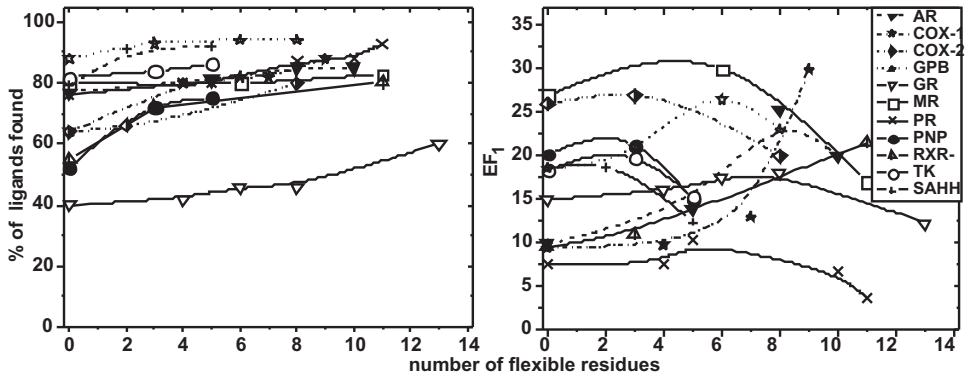


Figure 2. Variation of enrichment factors with the number of flexible residues employed in flexible docking.

present algorithm does not require empirical assumptions of possible soft spots of the receptor binding pocket. Instead, accuracy of the methed depends mainly on the number and variety of ligands known to be bound to the specific target. If only a limited number of known ligands is available, compounds with similar physical properties may be used to explore the active site and create a list of flexible residues.

References

1. H. Merlitz and W. Wenzel, *Comparison of of stochastic optimization methods for receptor-ligand docking* , J. Chem. Phys. Lett. **362**, 271-277, 2002.
2. W. Wenzel and K. Hamacher, *Stochastic tunneling approach for global optimization of complex potential energy landscapes*, Phys. Rew. Lett. **82**, 3003-3007, 1999.
3. W. L. Jorgensen and N. A. McDonald, *Development of an all-atomic force field for heterocycles properties of liquid pyridine and diazenes* , J. Mol. Struct. **424**, 145, 1997.
4. *Molecular Operating Environment (MOE) version 2003.02; Chemical Comuting Group Inc.:Monreal, Canada*, 2003.
5. G. M. Morris, D. S. Goodsell, R. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding energy function* , J. Comptut. Chem. **19**, 1639, 1998.
6. N. Huang, B. K. Shoichet, and J. J. Irwin *Benchmark sets for molecular docking*, J. Med. Chem. **49**, 6789-6801, 2006.
   `http://blaster.docking.org/dud/`.

Contributions to the Hydration Free Energies of Amino Acids

Gerhard König and Stefan Boresch

Department of Computational Biological Chemistry, University of Vienna,
Währingerstr. 17, A-1090 Vienna, Austria
E-mail: {gerhard, stefan}@mdy.univie.ac.at

Molecular solvation is a fundamental factor in biological processes, such as protein folding, receptor binding or enzymatic reactions. Currently, estimates of the hydrophobicity of amino acids are often derived from solvation (or transfer) free energies of side chain analogs. Such an approach implicitly assumes that contributions from the backbone and the side chain to the free energy of solvation are additive. However, it is well known that, in particular for polar amino acids, the properties of side chain analogs and amino acids can deviate significantly. Based on the relative hydration free energies of the amino acid pairs Ala-Ser, Val-Thr, Phe-Tyr, Val-Ala, Thr-Ser, Phe-Ala, and Tyr-Ser determined from molecular dynamics simulations, we quantitatively trace the molecular origin of these deviations to two effects, solvent exclusion and self-solvation. Solvent exclusion accounts for the reduction in solute-solvent interactions as one part of the solute occludes other parts of the solute, e.g., the presence of the backbone lowers the degree of direct interaction possible between the side chain and surrounding water. While solvent exclusion applies to polar and apolar amino acids alike, self-solvation is specific to polar amino acids and results from strong, directed intramolecular interactions between the polar functional groups of the side chains and polar moieties in the backbone, often through hydrogen bond formation. Thus, the contribution of self solvation to the solvation free energy is strongly conformation- and environment-dependent, and, therefore, the correct treatment thereof poses a challenge to applications involving solvation processes. Implications for the utility of hydrophobicity scales and connections to implicit solvent models are briefly discussed.

1 Introduction

Proteins, like most other biological macromolecules, function in aqueous solution. Therefore, one has to take into account the influence of solvent on the structure and thermodynamics of proteins, in order to understand their biological function. One fundamental principle for the description of this effect is hydrophobicity, which is often quantified by the partitioning of (model) compounds between water and an apolar medium.

Despite many successes, there are well documented problems associated with the use of hydrophobicity scales derived in this manner: First, very different transfer free energies were obtained depending on the solvent used for the apolar phase[1]. Second, in many approaches the raw data are obtained from side chain analogs, but are applied to the corresponding amino acids. This assumes that solvation free energies are additive, i.e. that the solvation free energy for the amino acid of interest is the sum of the solvation free energy of glycine (accounting for the contribution from the peptide backbone), and the solvation free energy of the side chain analog. However, as can be seen in Fig. 1a there are significant deviations ($\Delta\Delta\Delta A_{solv}$) from this assumed additivity relationship[2–5]. In the most extreme case (Ala–Ser), $\Delta\Delta\Delta A_{solv}$ is almost 5 kcal/mole. A high $\Delta\Delta\Delta A_{solv}$ can also be seen in other apolar-polar pairs (Val–Thr). For amino acid pairs of like polarity and relatively similar size (Ala–Val, Ser–Thr), the differences of approximately 1 kcal/mole

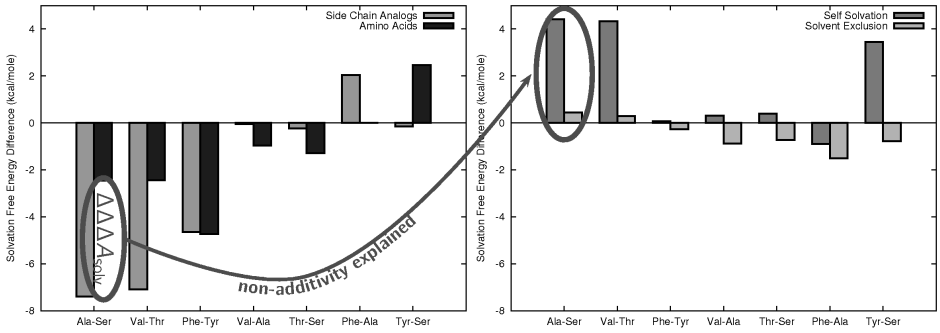


Figure 1. a.) Relative solvation free energy differences of selected amino acid pairs ($\Delta\Delta A^{AA}_{solv}$) and their corresponding side chain analogs($\Delta\Delta A^{SC}_{solv}$). The resulting difference between $\Delta\Delta A^{AA}_{solv}$ and $\Delta\Delta A^{SC}_{solv}$ ($\Delta\Delta\Delta A_{solv}$) is caused by interactions between the amino acid backbone and the side chain (the so-called "non-additivity"). The contributions of the two main effects causing non-additivity in amino acids (self-solvation and solvent exclusion) are depicted in part b.) on the right side of the figure.

between side chain analog and amino acid results are statistically significant, but much smaller than those obtained for the apolar–polar pairs. As the difference in size between two amino acids of similar polarity increases, so does the deviation from the respective side chain analog results (Ala–Phe, Ser–Tyr).

Several effects were assumed to be responsible for these deviations: First, one has to take into account the *solvent exclusion*, since the backbone reduces the solvent accessible surface area (SASA) of the side chain and vice versa. The second important concept has been referred to as *self-solvation*[3,4]. In apolar media, polar side chains interact with the polar groups in the peptide backbone, which effectively lowers the hydrophilicity of the side chain, as well as of the backbone.

The goal of the present study is to determine the molecular origin of any deviations from the additivity relationship, rather than verifying the mere existence of such differences. For this purpose, we concentrated on only a few amino acids which represent a broad range of distinct physicochemical properties (e.g. polarity and size). Furthermore, we (primarily) calculated relative rather than absolute hydration free energy differences.

2 Methods

Relative solvation free energy differences ($\Delta\Delta A_{solv}$) were calculated for selected pairs of N-acetyl-X-methylamide amino acids ($\Delta\Delta A_{solv}$ of Ala–Ser, Val–Thr, Phe–Tyr, Val–Ala, Thr–Ser, Phe–Ala, and Tyr–Ser) and pairs of the corresponding side chain analogs ($\Delta\Delta A^{SC}_{solv}$ of methane–methanol, propane–ethanol, toluene–*p*-cresol, propane–methane, ethanol–methanol, toluene–methane and *p*-cresol–methanol) by using thermodynamic integration and Non-Boltzmann Thermodynamic Integration[6]. The CHARMM22 all-atom protein force field[7] was used.

Gas phase free energy differences were obtained based on Langevin dynamics simulations at 300 K. All gas phase simulations had an overall length of 84 ns and were repeated at least five times. The solvent simulations lengths varied between 2.1 (for side chain analogs) and 42 ns (for full amino acids, using NBTI). Solvent simulations were repeated

at least three times.

3 Results and Discussion

To estimate contributions from solvent exclusion we computed solvation free energies for amino acids with the charges of all backbone atoms set to zero ($\Delta\Delta A_{solv}^{unch.\,BB}$), since the presence of the uncharged backbone prevents a complete solvation of the side chain. Thus, by taking the difference between $\Delta\Delta A_{solv}^{unch.\,BB}$ and $\Delta\Delta A_{solv}^{SC}$, one can estimate the free energy contribution from solvent exclusion by the backbone on the side chain.

Since the removal of either the backbone or side chain charges ($\Delta\Delta A_{solv}^{unch.\,SC}$), or all charges ($\Delta\Delta A_{solv}^{LJ}$) also leads to the extinction of all electrostatic interactions between side chain and backbone (thus also eliminating self-solvation), the combination of $\Delta\Delta A_{solv}$, $\Delta\Delta A_{solv}^{unch.\,BB}$, $\Delta\Delta A_{solv}^{unch.\,SC}$ and $\Delta\Delta A_{solv}^{LJ}$ can be used to estimate the self solvation.

Considering $\Delta\Delta\Delta A_{solv}$ in Fig. 1a and the self-solvation and solvent exclusion contributions in Fig. 1b, it is possible to distinguish two cases: First, the mutations involving apolar amino acids (Val–Ala, Phe–Ala): Here, the side chain analogs give acceptable approximations, if solvent exclusion is considered. Second, the mutations involving polar amino acids (Ala–Ser, Val–Thr, Tyr–Ser), where self-solvation plays a considerable role. However, there are mutations involving polar amino acids, but without a large self-solvation term (Phe–Tyr and Thr–Ser): In the case of Tyr, the polar groups are too distant from the backbone, and therefore there are no contributions from self-solvation. In Thr–Ser, on the other hand, the side chain hydroxyl groups are almost equidistant from the backbone (2.31 and 2.45 Å), leading to very similar self-solvation contributions, which cancel each other out. Our results imply that a correct treatment of both self-solvation and solvent-exclusion is fundamental for an accurate estimation of the solvation free energy of biomolecules (e.g. in implicit solvent models).

Acknowledgments

This work was funded by project P-19100 of the Austrian FWF.

References

1. P A Karplus, *Hydrophobicity regained*, Protein Sci., **6**, 1302, 1997.
2. J Chang, A M Lenhoff, and S I Sandler, *Solvation Free Energy of Amino Acids and Side-Chain Analogues*, J. Phys. Chem. B, **111**, 2098–2106, 2007.
3. L M Yunger and R D. Cramer III, *Measurement and Correlation of Partition-Coefficients of Polar Amino Acids*, Mol. Pharmacol., **20**, 602–608, 1981.
4. M A Roseman, *Hydrophilicity of polar amino-acid side-chains is markedly reduced by flanking peptide bonds*, J. Mol. Biol., **200**, 513–522, 1988.
5. W C Wimley and S H White, *Experimentally determined hydrophobicity scale for proteins at membrane interfaces*, Nat. Struct. Biol., **3**, 842–848, 1996.
6. M Leitgeb, C Schröder, and S Boresch, *Alchemical free energy calculations and multiple conformational substates*, J. Chem. Phys., **122**, 084109, 2005.
7. A D MacKerell et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*, J. Phys. Chem. B, **102**, 3586, 1998.

Conservation Analysis of Functional Important Residues of the Oxygen Evolving Mechanism Located in the D1 Subunit of Photosystem II

Eva-Maria Krammer and G. Matthias Ullmann

Structural Biology / Bioinformatics Group, University of Bayreuth,
95447 Bayreuth, Germany
E-mail: {*eva-maria.krammer, matthias.ullmann*}*@uni-bayreuth.de*

Key residues of the oxygen evolving center of Photosystem II are examined by a conservation analysis using a previously constructed profile Hidden Markov Model. The analysis revealed for some of the crucial residues a unexpected flexibility of the aminoacid character.

1 Introduction

In the evolution of life on earth, oxygen producing photosynthesis is of central importance. The oxygen evolving center is part of Photosystem II in plant chloroplasts and is located near to the luminal site in the reaction center formed by the subunits D1 and D2 of Photosystem II. Water binds to the manganese cluster of the reaction center. During the water splitting reaction, electron and protons are abstracted from water involving the nearby Tyrosine Y_Z (Tyr161 of the D1 subunit). After four oxidation steps, molecular oxygen is released. From Y_Z, protons are transferred via several protein residues to the lumen and the electrons are used to restore the oxidized special pair. It is known that in addition to Y_Z and the manganese cluster, a calcium ion and a chloride ion are needed for the oxygen formation, but the detailed mechanism of oxygen formation is still under debate. Even in the available crystal structures of Photosystem II with reasonable resolution, the exact organization of the manganese cluster is not clear.[1,2] However, several residues in the D1 subunit were suggested to coordinate the oxygen evolving center or to influence the oxygen formation due to interactions with Y_Z[3]. Namely Asp170, His332, Glu333, His337, His342 and the C-terminal of Ala344 are likely to coordinate manganese ions; Glu189 is suggested to coordinate the calcium ion and to be part of proton transfer pathway from Y_Z;[3] His190 abstracts a proton from Y_Z.[4]

In the here presented work, the conservation of functionally important residues is analyzed by sequence alignment using a previously constructed profile Hidden Markov Model.[5] Our analysis indicates that some of the functional important residues are not as strictly conserved as one might expect.

2 Material and Methods

Using a previously constructed profile Hidden Markov Model of the D1 subunit of Photosystem II and its ancestor subunit of bacterial reaction centers,[5] 226 D1 sequences were aligned using the HMMER software.[6] The so-constructed alignment was used for the conservation analysis.

Function	Residue	C (%)	ex	(%)	ex	(%)
Manganese coordination	Asp170	98.7	**Asn**	**0.9**	*His*	*0.4*
	His332	100.0				
	Glu333	100.0				
	His337	100.0				
	Asp342	100.0				
	Ala344	100.0				
Y_Z function	Tyr161	100.0				
	His190	100.0				
Calcium binding	Glu189	95.1	**Asp**	**3.5**	*Lys*	*1.3*

Table 1. Conservation analysis of residues of the D1 subunit suggested to be involved in oxygen formation. For each residue the conservation (C) and the exchange (ex) is given in %. Bold characters mark aminoacid exchanges, for which mutational studies in the cyanobacterium *Syncheocystis* sp. PCC 6803 that this mutant does not grow photoautotrophicallly. Italic character mark, aminoacid exchanges for which mutational studies in *Syncheocystis* sp. PCC 6803 showed, that the mutant still grows photoautotrophically.[3,7,8]

3 Results and Discussions

For all examined residues (see Table 1), a high conservation is observed, which is not surprising, since the sequences of the D1 subunits show an average sequence identity of about 85 %. Experimental evidence exists, that Asp170 is part of the high-affinity binding site for the first manganese ion which is first photo-oxidized during the light-driven assembly of the cluster.[9] Whether Asp170 also ligates the manganese ion, which is first photo-oxidized during the oxygen evolving reaction is still a matter of debate.[8–10] Because of its importance one would expect that especially Asp170 should be strictly conserved. Moreover, mutational studies showed that an exchange of Asp170 abolished oxygen evolution in most cases.[8] Nevertheless our study indicates that Asp170 is not strictly conserved in contrast to the other proposed manganese coordinating coordinating residues. We also observe an Asn and an His at the position 170. Asp170→His mutants are photoautothrophic, but in Asp170→Asn mutants oxygen evolution is nearly abolished.[8] It is possible that the exchange of Asp→Asn is only an artifact of the protein sequencing technique. But it might also be that although Asp170 is very important for the oxygen evolution, the protein can restore the function without this residue, maybe by second site mutations.

Tyrosine Y_Z (Tyr161) is crucial for the function of the Photosystem II reaction center, since through its radical state, it transports the electrons from the water (bound at the manganese cluster) back to the special pair and accepts also the protons during water splitting. Since it is assumed, that Y_Z is initially protonated, a base is needed near to Y_Z to abstract the proton of Y_Z. His190 has been suggested to be this base.[4] Both His190 and Tyr161 are strictly conserved in our analysis, which shows their crucial importance in the reaction mechanisms. Glu189 was proposed to accept a proton from His190 and thus to be part of a proton transport system. Moreover, it is suggested to coordinate the calcium ion.[2] Our study shows that Glu189 is not strictly conserved. The mutation Glu189→Asp in the cyanobacterium *Syncheocystis* sp. PCC 6803 led to organisms that could not grow photoautothrophically. However, we observe an Asp at the position of Glu189 in some of

the D1 sequences. Thus, it might be, that either another residue can take over the function of Glu189 or that organisms with Asp189 differ in their proton transfer mechanism.

4 Concluding Remarks

The functionally crucial residues such as Tyr161 and His160, as well as several suggested manganese coordinating residues (His332, Glu333, His337, Asp342 and Ala344) are strictly conserved. Surprisingly the residues Glu189 and Asp170, which are also thought to be very important for the function of Y_Z and the manganese cluster, are not conserved in the same extent. Moreover, they are also exchanged to aminoacids for which it is known that these mutations prevent oxygen production in the cyanobacterium *Syncheocystis*. It might be that although Asp170 and Glu189 are very important for the oxygen evolution that the protein function can be restored without these residues by second site mutations.

Acknowledgments

This work was supported by the Procope bilateral travel grant D/0502198. We thank Prof. P. Sebban for useful discussions.

References

1. B. Loll, J. Kern, W. Saenger, A. Zouni and J. Biesiadka, *Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II*, Nature **438**, 1040-1044, 2005.
2. J. Yano, J. Kern, K. Sauer, M. J. Latimer, Y. Pushkar, J. Biesiadka, B. Loll, W. Saenger, J. Messinger, A. Zouni adnd V. K. Yachandra *Where water is oxidized to dioxygen: structure of the photosynthetic Mn_4Ca cluster*, Science **314**, 821-825, 2006.
3. R. J. Debus, K. A. Campell, D. P. Pham, A.-M. A. Hays and R. D. Britt, *Glutamate 189 of the D1 polypeptide modulates the magnetic and redox properties of the manganese cluster and tyrosine Y_Z in photosystem II*, Biochemistry **39**, 6275-6287, 2000.
4. C. Tommos and G. T. Babcock *Proton and hydrogen currents in photosynthetic water oxidation*, Biochim. Biophys. Acta **1458**, 199-219, 2000.
5. E.-M. Krammer, P. Sebban and G. M. Ullmann, *submitted*.
6. S. R. Eddy *Profile hidden Markov models*, Bioinformatics **14**, 755-563, 1998.
7. H.-A. Chu, A. P. Nguyen and R. J. Debus *Amino acid residues that influence the binding of manganese or calcium to photosystem II. 1. The lumenal interhelical domains of the D1 polypeptide*, Biochemistry **34**, 5839-5858, 1995.
8. P. J. Nixon and B. A. Diner *Aspartate 170 of the Photosystem I1 Reaction Center Polypeptide D1 Is Involved in the Assembly of the Oxygen-Evolving Manganese Cluster?*, Biochemistry **31**, 942-948, 1992.
9. K. A. Campbell, D. A. Force, P. J. Nixon, F. Dole, B. A. Diner,and R. D. Britt *Dual-Mode EPR Detects the Initial Intermediate in Photoassembly of the Photosystem II Mn Cluster: The Influence of Amino Acid Residue 170 of the D1 Polypeptide on Mn Coordination*, J. Am. Chem. Soc. **122**, 3754 -3761, 2000.

10. R. J. Debus, M. A. Strickler, L. M. Walker, and W. Hillier *No Evidence from FTIR Difference Spectroscopy That Aspartate-170 of the D1 Polypeptide Ligates a Manganese Ion That Undergoes Oxidation during the S0 to S1, S1 to S2, or S2 to S3 Transitions in Photosystem II*, Biochemistry **44**, 1367 -1374, 2005.

Viral Membrane Proteins: Flexibility and Assembly

Jens Krüger and Wolfgang B. Fischer

Institute of Biophotonics, School of Biomedical Science and Engineering,
National Yang-Ming University, Taipei, Taiwan
E-mail: {jk, wfischer}@ym.edu.tw

Membrane bound and pore forming viral proteins like M2 from Influenza A, Vpu from HIV-1 or 3a from SARS-CoV show in their monomeric form high flexibility and adaptability in different lipid bilayer environments. Their conformational space under these conditions has been studied by ample molecular dynamic simulations.

The understanding of the abilities of the monomeric units is used to further explore the energy landscapes of the molecular assembly of multiple monomers. The newly developed protocol screens the full high dimensional search space leading to highly reliable pore models. After minor refinement e.g. with short molecular dynamics simulations they are suitable for the use in drug screening. Furthermore the evaluation of the energy landscapes allows drawing conclusions about the gating mechanism of the pores.

1 Introduction

The genome from different viruses encodes a series of membrane bound or attached proteins. These proteins fulfill a broad range of functions and are often essential in the virus reproductive life cycle (see also Patargias et al. in this proceedings p. 93). Some of these proteins include Vpu from HIV-1 which helps degrading the CD4 receptor and enhances the particle release (virion budding).[1,2] M2 from Influenza A which facilities the viral entry into the host cell via the endocytosis pathway.[3] 3a from SARS-CoV plays an important role during the virion release.[4] All of these proteins have in common that they assemble within a lipid environment to form multimeric homooligomers which function as pore or ion channel. Unfortunately all membrane proteins are barley crystallisable which makes it a challenging task to get atomistic x-ray data for this kind of proteins. Currently the best source for structural data are NMR studies, which are usually limited to monomeric forms and can only provide information about the assembled pore under special conditions.[3]

2 Computational Method

We herby present an approach to derive structural data for the assembly of viral membrane proteins within a two dimensional lipid bilayer environment. Starting point is to screen with experimental and computational methods the viruses protein sequence for putative transmembrane regions, which consequently could form a pore.

2.1 Monomer Molecular Dynamics

The transmembrane spanning parts are modeled as ideal helices and embedded into a phospholipid bilayer (POPC, but also DPPC, DDPC and DTPC). After stepwise minimization

they undergo multiple 10 ns GROMACS MD simulation with full pressure and temperature coupling. By applying a principal component analysis (PCA) the conformational space of the monomer is analyzed and an average structure is generated.[1] As the different proteins can develop significant kinks and bends this step is essential to derive a good starting structure for the following assembly.

2.2 Assembly Protocol

To sample the whole conformational space of a pore assembly the lipid environment is considered as two dimensional space in which the assembly takes place. Furthermore homooligomer pores are considered symmetrical towards their central pore axis. Monomers are placed around the central pore axis, while the following degrees of freedom are sampled in a systematic way. The distances between packed helices in transmembrane proteins usually show values around 10 Å. To cover weak and tight packing interhelical distances in the range from 8 to 12.5 Å are sampled. The expression 'angle' is used to describe the rotation of each monomer around its own helical axis. In the case of homooligomers there is only one value per conformation, as due to symmetry all monomers are oriented in the same way towards the central pore axis. In the case of heterooligomers it is necessary to sample multiple angles, one for each non-symmetrical monomer. In some cases like M2 from Influenza A it is possible to narrow the search space significantly, as it is known that His-37 and Trp-41 play an important role in the proton conductance through the pore. They have to face inwards which narrows the search by at least 2/3 to 120 °. The tilt describes the orientation of the helical axis towards the membrane normal. As membrane proteins can develop significant tilts up to 50 °,[1] it is also required to sample this dimension of the conformational space in a sufficient way. Finally the sidechains are optimized and the energy for each conformer is evaluated. The geometrical and energetic operations have been implemented with SVL in MOE (Chemical Computing Group, Montreal `http://www.chemcomp.com/`).

3 Pore Assembly

The quality of the pore models generated by the here presented protocol have been evaluated for M2 from Influenza A. The C_α-RMSD of the best generated pore compared to the established model 2H95 is 2.086.[3] This excellent agreement speaks for the strength of the approach. To illustrate the abilities of the above described assembly protocol detailed data for Vpu from HIV-1 is shown in Figure 1. In this case more than one minimum appears on the high-dimensional energy landscape. While some of the 5 resulting models may be excluded due to experimental evidence, some may represent alternative conformations, responsible for multiple conductance states.[2,5] To push the method to its limits it was attempted to generate a full pore model for 3a from SARS (data not shown). After assembling the monomeric unit with three membrane spanning parts, it was assembled to form a tetrameric pore.

4 Conclusion

Here we present a sophisticated approach to screen the conformational space of a protein on the basis of a forcefield which enables to scan the whole search space with an accept-

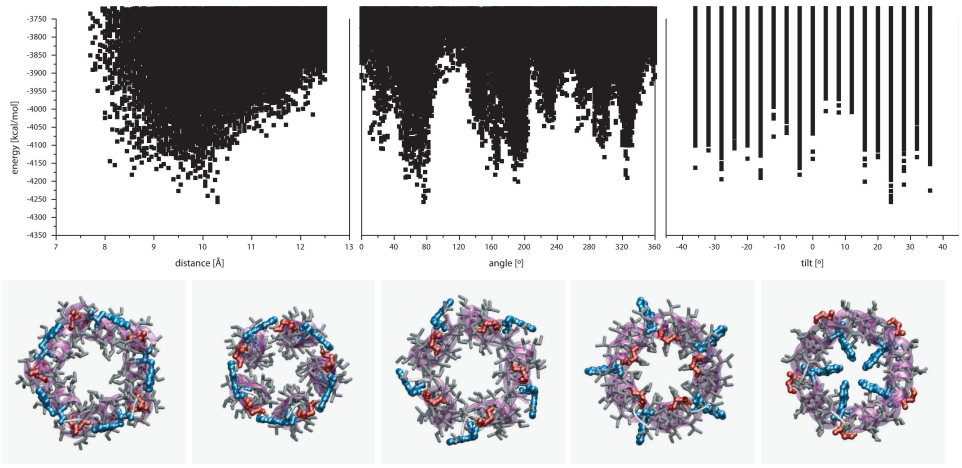


Figure 1. In the top row representative energyplots for the assembly of Vpu from HIV-1 are shown. The pore models which correspond to the minima in the plots are shown in the lower row. Trp-23 (blue) and Ser-24 (red) are highlighted. Both residues play an important role for the stability of the pore and the conductance of ions.

able resolution. By reasonable simplification and consideration of symmetry of the studied proteins a significant confinement of the search space can be made. This enables the resolvability of the search in an acceptable sampling time. The quality of the constructed structural models does not rank behind any experimental technique. In fact the careful optimization leads to more consistent models. When experimental results are taken into account e.g. in the form of a distance restraint, significant further confinements of the search space can be made. This boosts the sampling speed while at the same time an improved quality can be achieved.

Acknowledgments

Financial support from the National Yang-Ming University, the National Science Council Taiwan and the government of Taiwan (Aim for Top university plan) is acknowledged. We thank the PC^2 University of Paderborn, Germany for providing computing time.

References

1. J. Krüger and W. B. Fischer, *Exploring the conformational space of Vpu from HIV-1: a versatile adaptable protein*, J. Comp. Chem. , (in print), 2008.
2. Ed. W. B. Fischer, *Viral Membrane Proteins: Structure, Function and Drug Design*, Vol. 1, *Kluwer Academic / Plenum Publisher New York*, ISBN 0-306-48495, 2005.
3. J. Hu, T. Asbury, S. Achuthan, C. Li, R. Bertram, J. R. Quine, R. Fu and T. A. Cross, *Backbone Structure of the Amantadine-blocked Trans-Membrane Domain M2 Proton Channel form Influenza A Virus*, Biophys. J. **92**, 4335-4343, 2007.

4. W. Lu, B. J. Zheng, K. Xu, W. Schwarz, L. Du, C. K. L. Wong, J. Chen, S. Duan, V. Deubel and B. Sun, *Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release*, PNAS **103**, 12540-12545, 2006.
5. T. Mehnert, A. Routh, P. Judge, Y. M. Lam, D. Fischer, A. Watts and W. B. Fischer, *Biophysical characterisation of Vpu from HIV-1 suggests a channel-pore dualism.*, Proteins **70**, 1488-1497, 2008.

Hen Egg White Lysozyme Adsorption on a Mica Surface: A Fully Atomistic Molecular Dynamics Study

Karina Kubiak[1,2] **and Paul Mulheran**[1]

[1] Department of Chemical and Process Engineering, University of Strathclyde, Glasgow, Scotland
E-mail: karina.kubiak@strath.ac.uk

[2] Institute of Physics, N. Copernicus University, ul. Grudziadzka 5, 87-100 Torun, Poland

The interactions between proteins and solid surfaces are essential for a number of applications, such as functionalising biomaterials and for medical implants. The understanding of fundamental forces and processes involved in protein adsorption has a great importance in the construction of new, biocompatible materials. Our recent effort has been focussed on adsorption processes and protein dynamics on the surface, including protein cluster formation and cluster diffusion. Hen egg white lysozyme adsorption on a mica surface is an excellent model system for these investigations. Despite recent theoretical and numerical investigations, including Brownian Dynamics (BD) and Monte Carlo (MC) simulations, many atomistic details of lysozyme adsorption, cluster formation and protein/cluster(s) diffusion on the mica surface remains unknown.

Here, for the first time, we present results of fully atomistic, Molecular Dynamics (MD) simulations of hen egg white lysozyme (1iee.pdb) located in the neighbourhood of a mica surface. Protein adsorption is driven by electrostatic forces and so strongly depends on ionic strength. We have therefore examined two systems: solvated, neutral hen egg white with ionic strength equal 0.5 M and 0.02 M respectively. As a reference, a trajectory obtained for the isolated and solvated protein in ionic strength 0.5 M is also used. Careful analysis of four 20 ns trajectories provides an insight into early events during lysozyme adsorption on the mica surface, as well as influence of the surface and different ionic strength on the protein structure and stability.

1 Introduction

Interactions between proteins and solid surfaces are essential for a number of applications, such as functionalising biomaterials and for medical implants. The understanding of fundamental forces and processes involved in protein adsorption has a great importance in the construction of new, biocompatible materials. Our recent effort has been focused on adsorption process and protein dynamics on the surface, including protein cluster formation and cluster diffusion[1]. Hen egg white lysozyme (HEWL) adsorption on a mica surface is an excellent model system for these investigations. Despite recent theoretical and numerical investigations, including Brownian Dynamics[2,3], Monte Carlo simulations[4] and numerous Molecular Dynamics studies[5], many atomistic details of lysozyme adsorption, cluster formation and protein/cluster diffusion on the mica surface remains unknown.

Here, for the first time, we present results of fully atomistic Molecular Dynamics simulations of HEWL (1iee.pdb[6]) located in the neighbourhood of a mica surface. Protein adsorption onto a charged surface is driven mainly by electrostatic forces and so strongly depends on ionic strength. We have therefore examined two systems: solvated, neutral HEWL at pH=7 with ionic strength equal to 0.5 M and 0.02 M, respectively. As references, trajectories obtained for the isolated, solvated protein in these solutions are also

used. Careful analysis of these four 20 ns trajectories provides an insight into the early events during lysozyme adsorption on the mica surface, as well as the influence of the surface and different ionic strength on the protein structure and stability.

2 Methods

The crystal structure of HEWL (1iee.pdb[6]) was the starting structure of all our simulations, with all four disulphide bridges kept. The calculations were performed using the NAMD package[7]. The protein was placed in a rectangular box of water molecules (TIP3) that extend 8 Å from any protein atom. In the case of protein-surface system, a SiO_2 surface (mimicking a mica surface) with dimensions x=86.4 Å and y=92.8 Å was placed about 9 Å away from the closest HEWL side chain and 12 Å away from the HEWL backbone. The surface was created from silica and oxygen atoms, charged +1.11 e and -0.66 e respectively, located 1.6 Å away from each other, in a square array. The resulting surface charge density σ = -0.0217 e/Å^2 is almost equal to that of mica at pH=7 (see ref. 4). The HEWL-surface system was solvated in a water box that extends at least 20 Å from any protein atom. The protein or protein-surface systems were then neutralised by adding NaCl salt with ionic strength 0.5 M and 0.02 M. Most probable charge states at pH=7 were chosen for ionizable residues. These systems, composed from more than 15000 and 55000 atoms for HEWL and HEWL-surface respectively, were subject to 100 ps water equilibration, 10 000 steps of whole system minimisation, 30 ps heating to 300 K and 270 ps equilibration at this temperature. The production MD simulations were pursued for 20 ns at 300 K in the NVT ensemble. The integration step was 2 fs, and the SHAKE algorithm and periodic boundary conditions were used. The cutoff distance for both van der Waals and Coulomb interactions was 12 Å.

3 Results and Discussion

During the 20 ns trajectory for the HEWL-mica system with 0.5 M ionic strength solution, no attraction between the HEWL and surface was observed. Comparisons with the trajectory for the isolated HEWL show that the interactions and forces acting on the protein are not changed by the mica surface. Overall dynamical features such as protein mobility, including the mobility of the loops and changes in their conformation, show virtually the same interactions with water and salt ions. Based on this we can conclude that the electrostatic interactions are the main forces driving the adsorption process, and the solvent ionic strength 0.5 M effectively screens any electrostatic attraction between the negatively charged mica surface and the positively charged HEWL. Therefore the protein dynamics are not affected by the mica surface (data not shown) under these conditions.

A different picture emerges from the analysis of the 20 ns trajectory of the HEWL-mica system with 0.02 M ionic strength. In this case a strong attraction between the HEWL and mica surface is clearly visible, and the protein moves about 4-5 Å closer to the mica surface during the simulation. Graphical analysis reveals two simultaneous processes: the protein is moving as a whole towards the mica surface, whilst a conformational rearrangement occurs without losing secondary structure. This observation is supported by RMS and RMSF analyses. The RMS calculated for the HEWL-mica system stabilises after 1.4-2.0

ns at the 2.0-2.5 Å level, which means that the HEWL doesn't change its fold. The average fluctuations for the whole protein (Ca atoms) are 2.18 $Å^2$, for loop regions 2.47 $Å^2$ and for secondary structures 1.83 $Å^2$. From these RMSF results we can conclude that all helices and beta-sheets are maintained, and that conformational changes are accommodated only by loop and H-bonded turn rearrangement.

The HEWL conformation is more stable when the mica surface is absent; the RMS for the isolated HEWL reaches 1.5 Å after 20 ns (c.f 2.0-2.5 Å found above). Therefore the negatively charged mica surface placed close to the positively charged HEWL induces conformational changes accompanying the adsorption process.

The most important interactions between the mica surface and the HEWL are electrostatic attractions between the negatively charged surface and the positively charged Lys1, Arg14 and Arg128 residues. The distance between the surface and the Lys1 side chain is 11.3 Å at the beginning and 5.4 Å after 20 ns of simulation. In the case of the Arg14 side chain these distances are 11.3 Å and 7.9 Å after 20 ns, and for the Arg128 side chain 8.6 Å and 2.9 Å respectively.

The Lys1, Arg14 and Arg128 side chains form a triangle (side lengths 11 Å, 10 Å and 14 Å) whose conformation is better maintained when the mica surface is located close to the protein. All of these residues are strongly attracted by the surface and most probably the observed HEWL conformation changes are forced by this triangle maintaining its planar conformation.

4 Conclusions

We have found that the protein adsorption strongly depends on ionic strength, and conclude that the dominant interactions driving adsorption are the electrostatic interactions between the mica surface and the HEWL protein. Close proximity of the mica surface to the HEWL induces conformational rearrangement mostly confined to the loop regions, so that the HEWL secondary structure is maintained. The most important residues for the HEWL adsorption are Lys1, Arg14 and Arg128, whose rigid planar topology in close proximity to the mice surface appears to drive the loop conformational changes.

Acknowledgments

This work was supported by the UK Engineering and Physical Sciences Research Council through grant number EP/E012284.

References

1. P.A. Mulheran, D. Pellenc, R.A. Bennett, R.J. Green and M. Sperrin, *Mechanisms and dynamics of protein clustering on a solid surface*, Phys. Rev. Lett. **100**, 068102, 2008.
2. S. Ravichandran and J. Talbot, *Mobility of adsorbed proteins: a Brownian dynamics study*, Biophys. J. **78**, 110, 2000.

3. S. Ravichandran, J.D. Mandura and J. Talbot, *A Brownian dynamics study of the initial stages of hen egg-white lysozyme adsorption at a solid surface*, J. Phys. Chem. B. **105**, 3610, 2001.
4. F. Carlsson, E. Hyltner, T. Arnebrant, M. Malmsten and P. Linse, *Lysozyme adsorption to charged surfaces. A Monte Carlo study*, J. Phys. Chem. B. **108**, 9871, 2004.
5. A. Lerbret, F. Affouard, P. Bordat, A. Hedoux, Y. Guinet, M. Descamps, *Molecular dynamics simulations of lysozyme in water/sugar solutions*, Chem. Phys. **345**, 267, 2008.
6. C. Sauter, F. Otalora, J.A. Gavira, O. Vidal, R. Giege, J.M. Garcia-Ruiz, *Structure of tetragonal hen egg-white lysozyme at 0.94 A from crystals grown by the counter-diffusion method*, Acta Crystallogr. D. **57**, 1119, 2001.
7. J.C. Phillips, *et al*., *Scalable molecular dynamics with NAMD*, J. Comput. Chem. **26**, 1781, 2005.

Internal Dynamics of Ribosomal Elongation Factors G and Tu Studied with All-Atom and Coarse-Grained Molecular Dynamics

Katarzyna Kulczycka, Maciej Długosz, and Joanna Trylska

Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Żwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: {K.Kulczycka, M.Dlugosz, J.Trylska}@icm.edu.pl

Translation is modulated by various protein factors interacting with the ribosome. Elongation factor Tu (EF-Tu) delivers the aminoacyl-tRNA to the ribosomal A-site. After the peptide bond is formed, elongation factor G (EF-G) facilitates translocation of tRNAs to prepare the ribosome for the next catalytic cycle. The structure of EF-G resembles that of the complex between EF-Tu and aminoacyl-tRNA. We apply all-atom and coarse-grained molecular dynamics to search for common internal motions of these two molecules.

1 Introduction

Protein synthesis on the ribosome involves a number of protein factors that bind at its different functional sites. Our work focuses on two GTP-driven factors which share a common binding site: elongation factor G (EF-G) and elongation factor Tu (EF-Tu). EF-Tu transports the aminoacyl-tRNA (aa-tRNA) to the aminoacyl binding site (A-site) of the ribosome, in the form of the ternary complex EF-Tu-GTP-aa-tRNA. EF-G promotes translocation of the newly synthesized peptidyl-tRNA from the A-site to the peptidyl-tRNA binding site (P-site) together with its associated mRNA[1]. The structure of the EF-G resembles that of the complex between EF-Tu and aa-tRNA. This is an example of *molecular mimicry*[2]; a protein evolved so that its domains mimic the shape of a tRNA molecule. The N-terminal region of EF-G is homologous to EF-Tu, and the C-terminal region comprises a set of protein domains that adopted the shape of a tRNA [1]. We describe and compare internal dynamics of both factors, using full-atom and coarse-grained molecular dynamics (MD). Our aim was to check whether any similarity exists also in the dynamical behavior of the EF-G and EF-Tu-aa-tRNA complex.

2 Methods

Full-atom MD simulations were performed with the Amber9 package (`http://amber.scripps.edu/`). Structures of EF-G and EF-Tu-aa-tRNA were described according to the AMBER2003 force field. Mesoscopic model of solvent (GB^{OBC} model[3]) was used with dielectric constants set to 1 for solutes and 80 for water. Simulations were conducted at 150mM ionic strength. The temperature, set to 293K, was controlled with Andersen scheme[4]. Non-bonded interaction cutoff distance was set to 18Å. To analyze 20ns full-atom MD trajectories we applied principal component analysis[5] (PCA). Coarse-grained MD employed the Reduced Molecular Dynamics (RedMD)

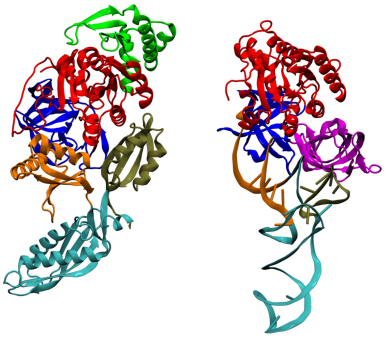


Figure 1. Left: the structure of EF-G (PDB code 1FNM[7]); red: domain I(G), green: insert (G'), blue: domain II, orange: domain III, cyan: domain IV, tan: domain V Right: the structure of EF-Tu:aa-tRNA complex (PDB code 1TTT[8]); red: domain I(G), blue: domain II, orange: the acceptor stem of aa-tRNA, cyan: the anticodon arm of aa-tRNA, tan: the T-arm of aa-tRNA. Domains I(G) and II of the EF-Tu and EF-G are homologous and common for GTPases[1].

package[6], a novel software which has been recently developed in our laboratory, with protein residues and RNA nucleotides represented as beads, interacting through harmonic (for neighboring) or Morse (for nonbonded) potentials[6]. Dynamics was simulated using Langevin equation (293K) and 1μs trajectory were generated for each molecule.

3 Results

Figure 2 presents correlation matrices derived from full-atom MD trajectories. For both molecules, motions of residues from distinct domains are correlated and the magnitude of this correlation is similar for homologous domains I and II. Strongest correlation occurs for

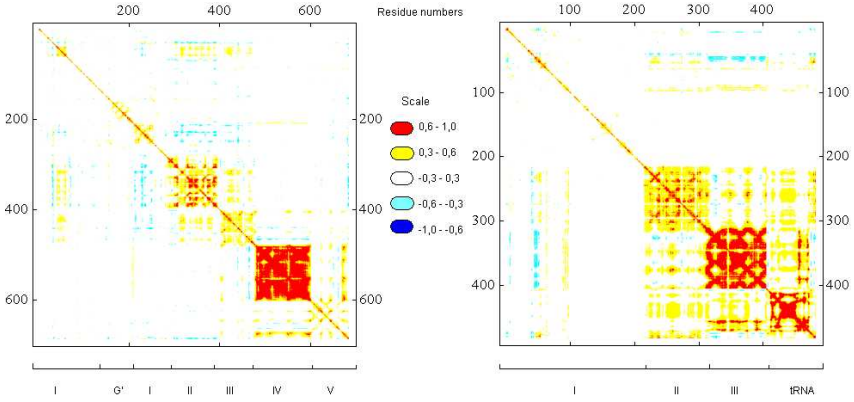


Figure 2. Correlation matrices: EF-G (left) and EF-Tu-aa-tRNA (right).

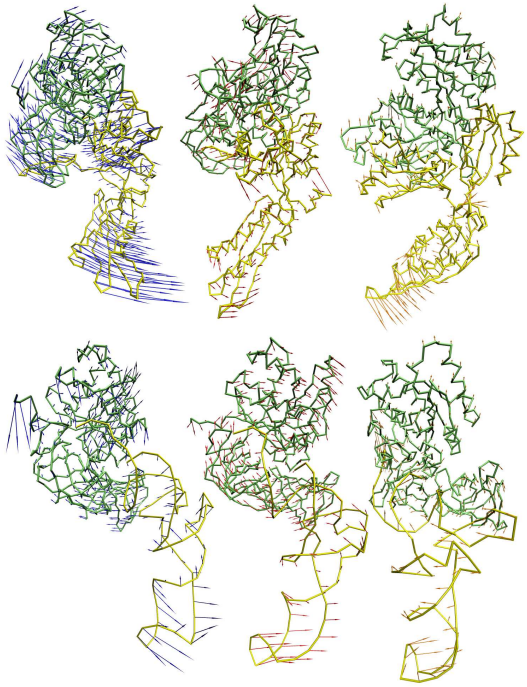


Figure 3. Graphical representation of collective motions in EF-G (top) and EF-Tu-aa-tRNA (bottom). Domains III, IV and V of EF-G and aa-tRNA are marked in yellow. Arrows show the directions of the first three eigenvectors of PCA analysis.

residues of domain IV of EF-G and domain III of EF-Tu-aa-tRNA complex. Three groups (the T arm, the acceptor stem and anticodon loop) are distinguishable in the nucleic part of the EF-Tu-aa-tRNA complex with strongly correlated movements of nucleotides within each group.

To characterize the most significant collective modes of motions we applied PCA to full-atom MD trajectories (Figure 3). The most dominant motions are those of domains III, IV and V of EF-G and those of aa-tRNA of EF-Tu-aa-tRNA complex. In both cases, the first two PCA principal components describe pendulum-like motions (in different planes) of the three EF-G domains and aa-tRNA, relative to the homologous domains I and II. Third PCA eigenvector describes pendulum-like motions of EF-G domains but for EF-Tu-aa-tRNA a stretching mode is also seen. We also compared the internal dynamics obtained with full-atom and coarse-grained MD. Figure 4 shows that root mean square fluctuations (RMSF) of C_α (EF-G) and C_α and P atoms (EF-Tu-aa-tRNA) observed in full-atom and coarse-grained simulations are of similar magnitude.

4 Conclusions

We compared internal dynamics of two structurally similar systems: protein EF-G and protein:RNA complex (EF-Tu-aa-tRNA) applying full-atom and coarse-grained MD. Our study demonstrated a certain degree of similarity in their dynamics, however, still more

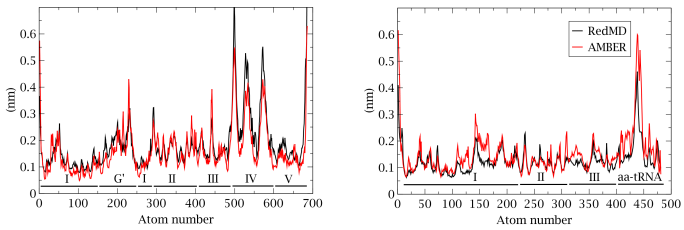


Figure 4. RMSF of C_α (EF-G, left) and C_α and P atoms (EF-Tu-aa-tRNA, right) derived from full-atom and coarse-grained MD simulations.

work is required to investigate its character[9].

Acknowledgments

We acknowledge support from University of Warsaw (115/30/ E-343/S/2007/ICM BST1255 and 115/30/E-343/BST1345/ICM2008), Polish Ministry of Science and Higher Education (3 T11F 005 30), Fogarty International Center (R03 TW07318) and Foundation for Polish Science. Simulations were performed at ICM University of Warsaw (G31-4).

References

1. A. Liljas. *Structural Aspects of Protein Synthesis*, World Scientific Publishing Co. Pte. Ltd., 2004.
2. A. T. Godkov, *Structure and function of the prokaryotic elongation factor G*, Mol. Biol. **35(4)**, 552–558, 2001.
3. A. Onufriev, D. Bashford, and D. A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized Born model*, Proteins: Struct. Funct. and Bioinf. **55**, 383-394, 2004.
4. T. A. Andrea, W. C. Swope, and H. C. Andersen, *The role of long ranged forces in determining the structure and properties of liquid water* J. Chem. Phys. **79**, 4576-84, 1983.
5. A. Amadei and A. B. M. Linssen and H. J. C. Berendsen, *Essential Dynamics of Proteins* Proteins: Struct. Funct. Genet. **17**, 412-425, 1993.
6. A. Górecki, M. Szypowski, B. Domagała, and J. Trylska, *RedMD - coarse-grained molecular dynamics package*, in preparation.
7. M. Laurberg, O. Kristensen, K. Martemyanov, A. T. Gudkov, I. Nagaev, D. Hughes, and A. Liljas, *Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site*, J. Mol. Biol. **303**, 593-603, 2000.
8. P. Nissen, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F.C. Clark, and J. Nyborg, *Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog*, Science **270**, 1464-1472, 1995.
9. K. Kulczycka, M. Długosz, J. Trylska, *Exploring internal dynamics of elongation factors G and Tu*, in preparation.

Efficient Molecular Docking of Drug Molecules into DNA Targets and their Enrichment by Cutting-Edge Technologies

S. A. Kumar[1,2], M. Petersen[2], M. G. B. Drew[1], M. M.-L. Grage[2], C. J. Cardin[1], and Y. Gan[1]

[1] School of Chemistry, University of Reading, Whiteknights, Reading RG6 6AD, Berkshire, UK
E-mail: scr04sa@rdg.ac.uk

[2] Nucleic Acids Center, Department of Physics and Chemistry,
University of Southern Denmark, DK-5230 Odense C, Denmark

A novel DNA mediated anti-cancer mechanism is aimed at stabilizing DNA G-Quadruplex structures. G-tetrads display high polymorphism and are found in the telomeric ends of chromosomes. In normal cells they exist as G-tetrads, producing knots that prevent the access of chromosomal DNA to the replication machinery. During cell division, these G-tetrad knots are resolved and the G-rich DNA is exposed to the replication machinery as single strands. Small molecules stabilizing G-Quadruplex DNA structures are thought of as potential novel anti-cancer agents. Several novel DACA analogues have been designed, synthesized and crystallized in reaction with the G-Quadruplex DNA. Regrettably they failed to diffract even synchrotron light. In the present investigation these analogues were studied for their binding interactions and stabilisation potential of a G-Quadruplex DNA $(TGGGGT)^4$ using molecular mechanics and combined QM/MM molecular docking techniques.

1 Introduction

The premise for the present study is that stable drug bound DACA analogue complexes with minimal bound free energy show potential to evolve as anti-cancer agents. A variaty of docking algorithms have been tested for docking a number of DACA analogues in the G-Quadruplex DNA. The resulting binding poses have been compared to the crystal structure of a similar drug molecule, daunomycin[1] (PDB ID:1O0K). It turned out that the results of spherical polar coordinate shape complimentarity based HEX[2] docking, and fragment growth based Glide-XP[3] (using the OPLS 2005 force field) resulted in realistic binding poses. The results were enriched by describing the polarization of the charge field in the drug - DNA interface by QM/MM single point calculations. The QM method was density functional theory (DFT) B3LYP and calculations were carried out wihth both 3-21G and 6-31G* basis sets through an iterative Quantum Polarized Docking Workflow[4]. In the following the results of the different molecular docking techniques will be discussed with an emphasis on the differences between force field based and combined QM/MM based predictions.

2 Motivation

Molecular docking is a means of computationally investigating ligand binding to a receptor in order to reduce the laboratory work, as well as justify chemical/physical explanations

for observed fenomena. Docking programs employ an empirical energy function to determine and optimize the interaction energy between the drug candidate and the active site. Structure-based docking methods automatically sample ligand conformations and protein-ligand interactions with a specified region of the protein surface. MM docking results are reported as the lowest energy/highest scoring pose(s) for each ligand, typically one pose for a large compound collection. Hybrid force field electronic structure docking algorithms divide the energy function into a quantum mechanical part describing the ligand, a force field based part describing (most of the) receptor and an interaction part taking care of the embedding of the two descriptions in the area where they interact. In the current case, we see the application of molecular docking as a trial and error process and seek to validate the predictions based on comparison with similar structures solved crystallographically or by NMR studies. Rigid docking turned out not to be succesful in producing the drug binding poses observed in the crystal structure. This served as a motivation to examine higher accuracy methods.

3 Materials and Methods

Different DACA analogues were sketched and minimized in gas phase using the UFF force field to prepare an ensemble of starting structures of drug molecules with no atomic clashes in their geometries. Several different docking algorithms were used at varying levels of accuracy, i.e. pure molecular mechanics and combined QM/MM methods. Of the algorithms tested, the HEX and the Schrodinger GLIDE methods produced good binding predictions which agreed with the binding clues learnt from a similar drug bound crystal structure. All docking experiments were conducted by blind docking, i.e. without specifying the binding site of the drug molecule in the receptor.

Glide XP is a force field based fragment docking method. QM/MM docking is achieved in combination with the IMPACT, Q-SITE and JAGUAR modules of Schrodinger suite, using primarily the OPLS 2005 force field. The Glide XP scoring function is designed to improve the results compared to the SP scoring function. This is achieved by adding terms to the function, the additional terms include: i) Coulomb energy of interacting atoms, ii) Van der Waal's energy of atoms, iii) Terms to favour binding interactions, iv) Terms to hinder binding interactions.

The terms that favour binding interactions include terms with parameters to evaluate the hydrophobic enclosures, hydrogen bonding possibilities between neutral-neutral H-bonding motifs, hydrogen bonding possibilities between charged-charged motifs, pi stacking interactions and pi-cation interactions. The terms that hinder binding interactions include terms with parameters to score desolvation in the binding site, and a term to calculate the inter-molecular strain energy based on proximity distances of ligand heavy atoms. On the other hand, the Hex docking algorithm is purely a structure complimentarity matching algorithm based on polar spherical coordinates. The molecular surface is represented by several radial expansions of spherical harmonic basis functions. Surfaces are generated for both receptor and ligand molecules and the shape complimentarity is scored to arrive at binding predictions.

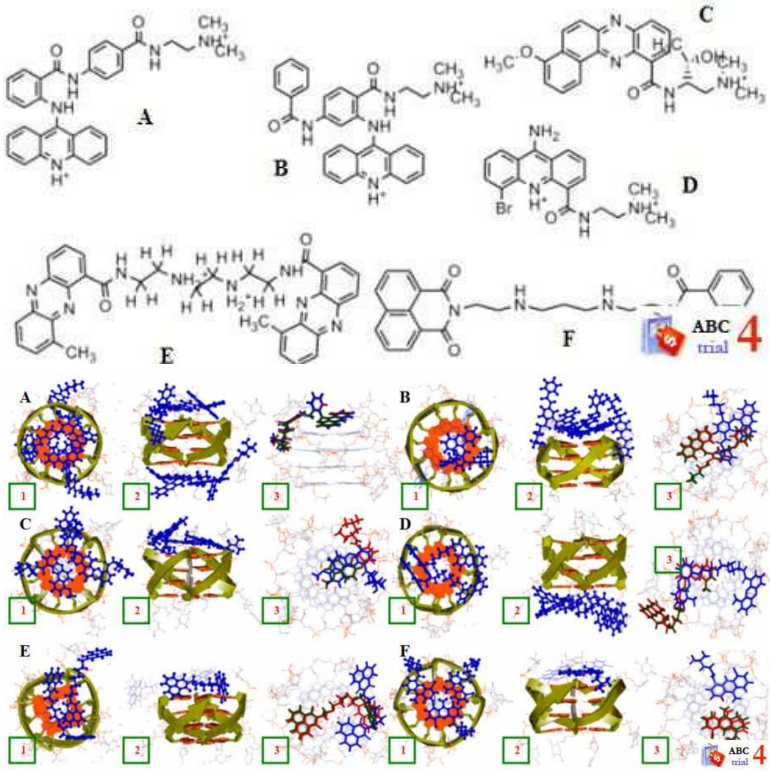


Figure 1. A = Drug 1, B = Drug 2, C = Drug 3, D = Drug 4, E = Drug 5, F = Drug 6. 1 & 2 = Hex docking poses, 3 = Glide docking poses. Colour Codes: Blue = Glide XP OPLS 2005, Red = QM/MM B3LYP with 3-21G basis sets, Green = QM/MM with 6-31G* basis sets.

4 Results and Discussions

There is an overall agreement between the results of Hex and Glide XP docking predictions $fig.1$, although minor variations exist in i.e. side chain conformations etc. In general Glide XP placed the conformations better than the Hex poses which might be due to the bias of the pure shape complimentarity principle of Hex. This resulst in the deviations of chemical geomtry which are implied better in various force fields.

The important binding features deduced from the crystal structure are:

1) Planar aromatic rings formed stacking interactions with the DNA bases, 2) Flexible side chains were favored in the electrostatically stable groove regions of the DNA, 3) The G-Quadruplex terminal surface was enough only to accommodate two molecules of daunomycin in fully stacked manner.

Daunomycin is made up of a planar four-ring chromophore and a sugar moiety, while the DACA chromophore is a planar three-ring system with a more flexible side chain and no sugar moiety. As the DACA analogues contain aliphatic side chains in contrast to the heterocyclic sugar moiety in daunomycin, a more compact packing of the sidechains in the

DNA groove regions is expected.

DACA analogues; drug1, drug3 and drug4 are more simple structures containing comparatively shorter length aliphatic side chains connected to the aromatic ring system and thus the Hex and Glide predictions turned out to agree well. In the case of drug2, an aromatic ring branches out from the planar pharmacophore, which is further bifurcated with another aromatic ring on one side and an aliphatic side chain on the other side. Thus drug2 is quite complex in nature, something which might require knowledge of the physical/chemical parameters for a better prediction. The Hex binding pose exhibits an electrostatistically less favorable state where the bifurcated groups are dumped in the same region. This is in opposition to the Glide output in which the bifurcation maintains the placement of the aromatic and the aliphatic side chain bulks on opposite sides. In the current case, it can also be suggested that a combination of geometrical and force field methods can lead to a robust prediction mechanism.

DACA analogues; drug5 and drug6 are structures which can typically be considered as tougher molecules for docking. The complexity is due to their long, highly flexible aliphatic linkers. It evinces that approximations in force field parameters and/or negligence of force field parameters lead to highly deviating and less favorable binding predictions when compared to predictions made by combined QM/MM models. The inclusion of QM derived charge polarization clearly has exhibited a better performance that agrees with hints taken from the crystal structure. The Hex predictions place the linkers passing through the terminal surface of G-Quadruplex with the chromophores kept away from stacking interactions. Glide XP heavily under-performed the drug5 binding pose with many fewer interactions between the drug and DNA, which is rectified by redocking with included QM polarized charges.

In conclusion the charge polarization plays a very important role in drug - receptor recognition and binding and can be included by combining traditional MM and QM methods. Furthermore our experiments show that simple algorithms considering mainly the shape complimentarity perform nearly as well as the force field parameters based molecular docking methods. In overall, molecular docking can be better achieved by combining several different methods which utilizes a combination of chemical/physical information, shape complimentarity and induced effects in the drug-receptor interface.

Acknowledgments

We are very thankful to all our academic and industrial collaborators. Special thanks to Nucleic Acids Center at SDU, MC FP6 Project: Nucleic Acid Based Drug Design and the Danish Center for Scientific Computing.

References

1. Clark, G. R., Pytel, P. D., Squire, C. J., Neidle, S., J. Am. Chem. Soc., **125**, 4066–4067, 2003.
2. Ritchie, D. W., Kemp, G. J. L., J. Comp. Chem. **20(4)**, 383–395, 1999.
3. GLIDE, version 4.5, (Schrodinger, LLC: Portland, OR, 2000).
4. Cho, A. E., Gullar, V., Berne, B. J., Friesner, R., J. Comp. Chem. **26**, 915–931, 2005.

Applications of a Novel Biasing Potential to Study DNA Translocation and DNA Base Flipping

Sean M. Law and Michael Feig

Department of Biochemistry & Molecular Biology, Michigan State University,
East Lansing, Michigan, 48824-1319, USA
E-mail: {*slaw, feig*}*@msu.edu*

DNA transcription, replication, and damage repair usually involve DNA-protein interactions and structural distortion of the DNA duplex by various enzymes. For example, during DNA metabolism, DNA helicases have been shown to separate duplex DNA into individual strands by translocating along single stranded DNA (ssDNA) while hydrolyzing ATP [1,2]. Alternatively, various enzymes employ a base-flipping mechanism to tackle DNA repair [3]. Experimental studies have demonstrated that DNA translocation and flipping of DNA base pairs typically occurs on the millisecond or longer timescale [4,5]. However, current computational methods are limited to the nanosecond timescale. Thus, external restraints are often employed to enhance sampling in these low probability regions of phase space. While flipping of individual bases using different restraints has been well established [3], computational studies related to DNA translocation with respect to proteins is, to the best of our knowledge, slowly emerging [6]. In this study, umbrella sampling with a novel center-of-mass projection onto a predefined path reaction coordinate was utilized to study DNA translocation in the context of the transcription factor, E2F-DP, protein-DNA complex using implicit solvent.

1 Introduction

Protein-DNA interactions involving DNA base flipping and DNA translocation are vital for the proper functioning and survival of a cell. However, studying the mechanisms and energetics associated with these processes using straight molecular dynamics simulations are often unfeasible within the available computer time. The novel biasing potential presented below was used to study DNA translocation (in the presence of a DNA-binding protein) and is currently being applied to DNA base flipping (in the absence of a protein) though its function can be extended beyond these applications.

2 Development of New Biasing Potential

As a first step to understanding DNA translocation (and DNA base flipping), a new biasing potential, U_{res}, was developed and can be described as the projection of one or more center of mass (COM) onto a well understood path reaction coordinate. This harmonic potential (added to CHARMM [7])

$$U_{res} = k(t' - t_{initial} - t_0)^2 \tag{1}$$

can be expressed in terms of its initial COM projection onto the path $t_{initial}$, its COM projection onto the path at a given timestep t', its equilibrium COM projection value t_0 (relative to $t_{initial}$), and its force constant k. To elaborate, for a given well-behaved path

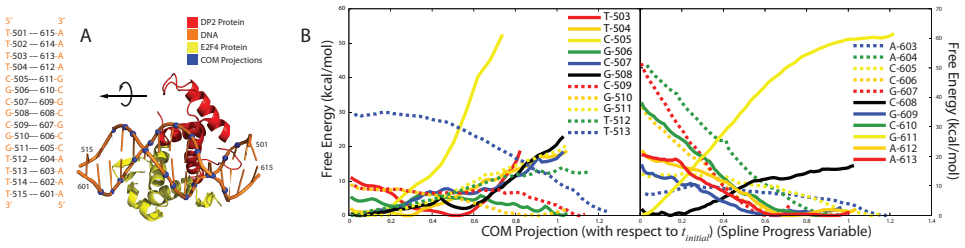


Figure 1. A. COM projection of select nucleotides onto a defined path. The translocation direction is shown in black arrows. B. Free energy profiles of each individual nucleotide. The units along the horizontal axis roughly corresponds to 3.8 angstroms (the rise in DNA).

represented by a set of points, cubic spline interpolation is used to reconstruct a piecewise smooth curve and the COM of interest is projected onto this curve. Since the projection of any point, $P(x, y, z)$, onto a cubic spline involves solving the roots of a quintic equation (which has no closed form solution), the method was simplified by projecting the COM onto a tangent line (with the assumption that this tangent is sufficiently close to the real tangent and remains constant within a given simulation window). Once the projection is approximated the umbrella potential is applied and is repeated for subsequent simulation windows by modifying t_0 at the start of each window.

3 Translocation of E2F-DP-DNA Complex

To study DNA translocation, the DNA-bound crystal structure of a heterodimeric transcription factor, E2F-DP, (PDBID: 1CF7[9]) was used where the path reaction coordinate was chosen as the DNA backbone. With the exception of the two terminal nucleotides from each DNA strand, the COM for each nucleotide was projected onto its own DNA backbone (Fig. 1A). After equilibration, the molecular dynamics simulations were carried out at 300 K using GBMV implicit solvent and free energy profiles were generated using the weighted histogram analysis method (WHAM) [8]. To facilitate DNA translocation by one base pair, all values of t_0 were uniformly adjusted over the course of 20 simulation windows away from DP and closer towards E2F, carrying out 50 ps of equilibration and at least 50 ps of sampling for each umbrella (while the path reaction coordinate remains unchanged).

The results obtained from WHAM analysis reveal important interactions between the amino terminus of the E2F protein and the DNA. With the exception of Cyt505 (and its hydrogen bonding partner, Gua611), the free energy profiles for each individual base (Fig. 1C) show general fluctuations in the relative free energies which can be associated with the breaking of hydrogen bonds between the protein-DNA complex. However, as Cyt505 moves along the reaction coordinate it encounters a key protein residue, Arg17 (from the amino-terminal of E2F), that is situated deep into the minor groove. According to Jordan et al, this conserved arginine appears to be important as its deletion abolishes DNA binding though its structural role is unclear [10]. Our findings suggest that once the E2F-DP dimer recognizes and binds to the central CGCGCG sequence, Arg17 plays an integral part in hindering DNA translocation.

4 DNA Base Flipping

For DNA base flipping, the COM of the central cytosine base (in GTCAGCGCATGG) was projected onto the path reaction coordinate which was chosen as the perimeter of a circle lying in the plane of the nitrogenous base and centered about the C3' atom of the residue of interest. To induce base opening, t_0 was adjusted over the course of 160 simulation windows towards the major and minor groove, respectively, carrying out 50 ps of equilibration and 200 ps of sampling for each individual window. The results to date are in good structural agreement with the current knowledge [3].

5 Acknowledgments

The authors would like to acknowledge financial support from NSF CAREER grant 0447799 and a QBMI Fellowship at Michigan State University (awarded to S.L) as well as access to computational resources at the High Performance Computing Center at Michigan State University.

References

1. Caruthers, J. M. and McKay, D. B., *Helicase structure and mechanism*, Current Opinion in Structural Biology **12**, 1 (123-133), 2002.
2. Saha, A. and Wittmeyer, J. and Cairns, B. R., *Chromatin remodelling: the industrial revolution of DNA around histones*, Nature Reviews Molecular Cell Biology **7**, 6 (437-447), 2006.
3. Priyakumar, U. D. and MacKerell, A. D., Jr., *Computational approaches for investigating base flipping in oligonucleotides*, Chem Rev **106**, 2 (489-505), 2006.
4. Dillingham, M. S. and Wigley, D. B. and Webb, M. R. *Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed*, Biochemistry **39**, 1 (205-12), 2000.
5. Gueron, M. and Leroy, J. L., *Studies of base pair kinetics by NMR measurement of proton exchange*, Methods Enzymol **261**, 383-413, 1995.
6. Yu, J. and Ha, T. and Schulten, K., *How directional translocation is regulated in a DNA helicase motor*, Biophys J **93**, 11 (3783-97), 2007.
7. Kumar, S. and Bouzida, D. and Swendsen, R. H. and Kollman, P. A. and Rosenberg, J. M., *The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method*, Journal of Computational Chemistry **13**, 8 (1011-1021), 1992.
8. Brooks, B. R., Bruccoleri, R. D., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations* Journal of Computational Chemistry **4**, 187-217, 1983.
9. Zheng, N. and Fraenkel, E. ad Pabo, C. O. and Pavletich, N. P., *Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP*, Genes & Development **13**, 6 (666-674), 1999.
10. Jordan, K., Haas, A., Logan, T., and Hall, D., *Detailed analysis of the basic domain of the E2F1 transcription factor indicate that it is unique among bHLH proteins*, Oncogene **9**, 1177-1185, 1994.

Characterization of the Quinolone-Gyrase-Interaction Using Docking, Molecular-Dynamics and Site-Directed Mutagenesis

Jörn Lenz[1,2], **Thomas Lemcke**[2], **Peter Heisig**[2], **and Andrew Torda**[1]

[1] Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany
E-mail: {jlenz, torda}@zbh.uni-hamburg.de

[2] School of Pharmacy, University of Hamburg, Bundesstrasse 45, 20146 Hamburg, Germany
E-mail: {lemcke, heisig}@chemie.uni-hamburg.de

Fluoroquinolones are an important class of anti-bacterials, but as with many anti-infectives, drug-resistance is an increasing problem. Fluoroquinolones inhibit DNA-gyrase, an enzyme, which is able to alter the topology of DNA. However, there is only little structural information. In the literature, there are two rather different proposals for the binding mode of quinolones. Via docking calculations and MD simulations, we find both binding possibilities and consider the evidence for each. Work is now underway to test our models with site directed mutagenesis techniques.

1 Introduction

Despite modern antibiotics, infectious diseases are responsible for nearly one third of human deaths worldwide, and bacterial resistance is still an urgent problem. The class of fluoroquinolone(FQ)-antibiotics is a good example. They offer a broad spectrum of activity, good pharmakokinetic properties and are relatively cheap to produce. However, FQ resistant bacteria are wide spread.

The functional target and biochemical action of FQs is known. The drugs inhibit the A_2B_2 DNA-gyrase tetramer, an enzyme which is able to alter the topology of DNA by transient cleavage[1]. This is performed by an esterification of GyrA-Tyr_{122}-OH of the enyzme to a $5'$-phosphate of the DNA. FQs appear to inhibit the religation of the DNA by stabilizing the cleaved form, ultimately resulting in bacterial cell-death[2]. Furthermore, studies of natural and in-vitro mutants offer clues as to which residues are involved in drug-resistance[3]. Unfortunately, the exact molecular action remains largely unknown. Clearly, understanding the drug-enzyme-DNA interaction in molecular terms could be the basis for the development of new FQ-derivatives refractory to resistance. Due to the lack of complete structural information for the tetramer in complex with DNA and FQ, we have been trying to build molecular models using a combination of docking methods and molecular dynamics (MD) simulations.

2 Methods

2.1 Protein-DNA-Docking and Filtering

The mode of gyrase action can roughly be divided into three individual steps. First, the DNA approaches the protein. Secondly, a phosphate (DNA) - tyrosine (protein) ester is

formed. Finally, the DNA gap is resealed by a trans-esterification and subsequent DNA release. How the protein and the DNA find each other can be regarded as a docking problem. Hence, the first step was to generate many possible DNA-enzyme conformations using a standard protein-DNA docking tool[4]. Results are filtered according to orientations of the DNA 5′-end w.r.t. to the protein's active site Tyr_{122}, distances between GyrA-Tyr_{122}-OH and appropriate DNA cleavage sites (5′-↓GRYC-3′)[5], as well as electrostatic surface potentials.

2.2 Molecular Dynamics Simulations and Molecular Docking

Two successive sets of MD simulations were performed to find plausible and stable models of the protein-DNA complex. In the first set, DNA and protein were not linked while in the second set, a covalent bond between GyrA-Tyr_{122}-OH and the DNA 5′-end was introduced. In between, implausible results were discarded. MD simulations, including 70.000 TIP4P water molecules, were run up to 2ns using GROMACS[6].

Subsequently, remaining complexes were used for molecular docking of eleven different quinolones using AutoDock4[7]. For each of the drugs, 250 different conformations and their binding modes within a box were sampled. The box was defined to contain the potential binding pocket between Ser_{83} and Asp_{87} in the QRDR (quinolone resistance determining region) of GyrA (residues 67-106) as well as the four overhanging bases of the cleaved dsDNA.

3 Results and Discussion

The first docking session (protein-DNA) yielded $>$1000 candidates of which 24 were selected for the initial MD simulations (section 2). Of these, two had distances between GyrA-Tyr_{122}-OH and the DNA 5′-end which allowed the DNA-protein ester to be introduced.

We can distinguish these complexes by the orientation of the nucleic acid w.r.t. the α_4 helix of the DNA binding HTH motif present in the QRDR. The first complex places α_4 in the major groove of the DNA as proposed by Liddington[1], whereas the second complex supports the model of Lapogonov[8] with α_4 in the minor groove of DNA. For each drug molecule in both complexes, a conformational cluster analysis was performed which was based on positional RMSD of corresponding drug atoms.

For our Lapogonov-like complex, we found a noticeable difference between the clusters containing the highest scoring FQ conformations. Although the drug molecules were placed within the same binding cavity, they appear to be rotated by $\sim$180° as shown in fig. 1. Amongst all sampled conformations, these drug arrangements were found most frequently without showing noticeable deviations with respect to their quantities. Our model indicates that the ligands might be spatially fixed with DNA and enzyme by two main anchors. On the one side, the N-containing heterocyclic substituent at C7 could be able to interact with the phosphate backbone of guanine or cytosine residues of the DNA. However, in its protonated form, this heterocycle might also interact with backbone oxygens of Ser_{83} and Ala_{84}. On the other side, the carboxyl moeity at C3 of the FQs might form a salt-bridge with the free amino-group of guanine. Moreover, our results could also allow the

presence of Mg^{2+} ions[9], which might fill the gap (7.5-8.6 Å) between the FQ's carboxyl- and keto-group, respectively, and the phosphate backbone of the DNA.

For the Liddington-like model, the highest scoring FQ conformations did not fit into the proposed binding cavity. However, as fig. 2 shows, lower scoring dockings are found for each FQ without clear preferences for a certain conformation. Anyhow, all dockings are in agreement with the common literature since they satisfy the intercalating nature of the quinolones.

This theoretical data enabled us to identify amino acids which are currently used to test the validity of our models with site directed mutagenesis.

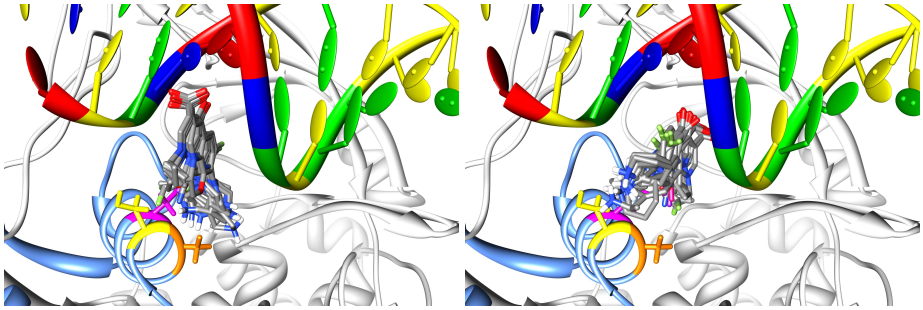


Figure 1. *Complex of DNA linked to GyrA-Tyr$_{122}$ and docked quinolones: COOH-moeities of ligands point out of (left) and into plane (right); enzyme shown in grey, QRDR in lightblue, Ser$_{83}$, Ala$_{84}$ and Asp$_{87}$ in α_4 helix in yellow, orange and magenta, resp.; G in green, C in yellow, T in blue and A in red.*

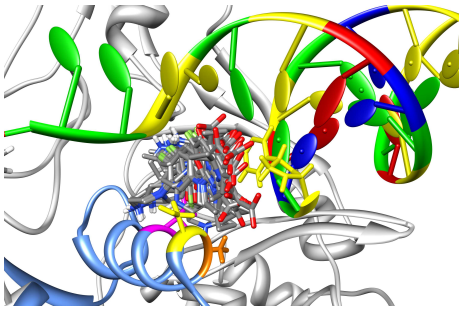


Figure 2. *Liddington-like complex of DNA linked to GyrA-Tyr$_{122}$ and docked quinolones; color coding as in fig. 1; pictures were built using UCSF Chimera[10].*

Acknowledgments

This work is generously supported by the *Jürgen Manchot Stiftung, Germany* and a grant of the *University of Hamburg, Germany*.

References

1. J H Cabral, A P Jackson, C V Smith, N Shikotra, A Maxwell, and R C Liddington, *Crystal structure of the breakage-reunion domain of DNA gyrase*, Nature, **388**, no. 6645, 903–906, 1997.
2. W A Goss, W H Deitz, and T M Cook, *Mechanism of action of nalidixic acid on Eschrichia coli - Inhibition of deoxyribonucleic acid synthesis*, J Bacteriol, **89**, 1068–1074, 1965.
3. H Yoshida, M Bogaki, M Nakamura, and S Nakamura, *Quinolone resistance-determining region in the DNA gyrase gyrA gene of Escherichia coli*, Antimicrob Agents Chemother, **34**, no. 6, 1271–1272, 1990.
4. D Schneidman-Duhovny, Y Inbar, R Nussinov, and H J Wolfson, *PatchDock and SymmDock: servers for rigid and symmetric docking*, Nucleic Acids Res, **33**, no. Web Server issue, W363–7, 2005.
5. C G Noble, F M Barnard, and A Maxwell, *Quinolone-DNA interaction: sequence-dependent binding to single-stranded DNA reflects the interaction within the gyrase-DNA complex*, Antimicrob Agents Chemother, **47**, no. 3, 854–862, 2003.
6. D Van Der Spoel, E Lindahl, B Hess, G Groenhof, A E Mark, and H J C Berendsen, *GROMACS: fast, flexible, and free.*, J Comput Chem, **26**, no. 16, 1701–1718, 2005.
7. G M Morris, D S Goodsell, R Huey, W Hart, R Belew, and A J Olson, *Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function*, J Comput Chem, **19**, 1639–1662, 1998.
8. I Laponogov, D A Veselkov, M K Sohi, X Pan, A Achari, C Yang, J Ferrara, L M Fisher, and M Sanderson, *Breakage-Reunion Domain of Streptococcus pneumoniae Topoisomerase IV: Crystal Structure of a Gram-Positive Quinolone Target*, PLoS ONE, **2**, e301, 2007.
9. G Palu, S Valisena, G Ciarrocchi, B Gatto, and M Palumbo, *Quinolone binding to DNA is mediated by magnesium ions.*, Proc Natl Acad Sci U S A, **89**, no. 20, 9671–9675, 1992.
10. E Pettersen, T Goddard, C Huang, G Couch, D Greenblatt, E Meng, and T Ferrin, *UCSF Chimera–a visualization system for exploratory research and analysis*, J Comput Chem, **25**, no. 13, 1605–1612, 2004.

Classification of Kinases: A Fast, Automated Structure-Based Approach

Jörn Lenz[1,2], **Thomas Margraf**[1], **Thomas Lemcke**[2], **and Andrew Torda**[1]

[1] Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany
E-mail: {jlenz, margraf, torda}@zbh.uni-hamburg.de

[2] School of Pharmacy, University of Hamburg, Bundesstrasse 45, 20146 Hamburg, Germany
E-mail: lemcke@chemie.uni-hamburg.de

We have been developing phylogenetic methods based on protein structures rather than sequence and fast enough to be applied to large families. A good example of a large family are kinases. Often, one believes in evolutionary relationships based on protein function, but one cannot see the relationships because the sequences have diverged so far. This is certainly the case with protein kinases. They are found in most forms of life, but with a tremendous spread of sequences and even differences in function. They are an ideal candidate for our approach since there are hundreds of known structures. Phylogenetic trees can be built automatically and they even map rather well to the biochemical annotations which were determined manually.

1 Introduction

Historically, phylogeny of large protein families has been based on sequence information. We have been developing methods, which are based on protein structure, but are still fast enough to be applied to large numbers of proteins. Here we consider the example of kinases, which are able to alter the activity of enzymes or other molecules by covalently attaching phosphate groups. This strategy usually denotes a response to chemical signals with some persistence, depending on reversibility and degradation mechanisms.

Kinases are central components in signal transduction networks and can be found in nearly all regulatory and metabolic processes in eukaryotes[1] and also many prokaryotes[2]. They play a major role in cell growth, division and controlled cell death, as well as in hormone response. Changes in human kinase activity can cause erroneous phosphorylation and trigger severe ailments such as cancer, diabetes or neurodegeneration[3]. Thus, kinases are suitable targets for the treatment of such diseases[4].

Understanding their evolution could help to explain the specific functions of individual kinases. It could assist in decoding signalling events and the emergence of pathologic biochemical processes. This might contribute to a more detailed insight of drug selectivity and drug cross reactivity and thus to the development of more effective drugs. It may also aid the selection of kinases used for drug screenings.

There is almost no significant sequence similarity between the more distant kinases, so it is difficult to build reliable sequence alignments. This superfamily, however, has been popular amongst crystallographers, so there is a wealth of solved structures. This makes it an ideal candidate for a structure-based phylogeny.

The only similar project in the literature was based on 31 kinase structures and required human intervention to construct a phylogenetic tree[5]. Here, we show how one can use many hundreds of structures to build a phylogeny completely automatically.

2 Methods

The list of kinase structures was assembled from successive structure searches[6]. Subsequently, a multiple alignment of those structures was computed using HANSWURST[7]. The guide trees, superimposed structures, and derived sequence alignments are analysed below. A more detailed description of the methods used is given by Margraf[7] and Lenz[8].

3 Results and Discussion

In this section we present an excerpt of the results of the Neighbour Joining clustering on RMSD values of pairwise superpositions (fig. 1). The multiple structure alignment and the implied multiple sequence alignment of the CMGC members show that the conserved features of kinases are appropriately superimposed. The most significant deviation from functional classification relates to the AGC kinases and the TK group. Firstly, 1h1wA is not clustered with any other member of the AGC group. Secondly, the G-protein coupled receptor kinase 1omwA is clustered with 1muoA from the "other" group. The superposition reveals that both structures share important features such as α-helix B[5] and could be structurally aligned. Thirdly, the AGC kinases 1cdkA and 1o6lA are clustered with 1phkA. The structural superpositions indicate that none of the alignments are unreasonable. Concerning the TK group members, only three of five were grouped together. However, they are interspersed with a TKL group kinase (fig. 1). The remaining two kinases of this group are clustered, but their distance to the CAMK group is closer than the distance to the other TK members. Another result of this work is the unusual placement of the TGFβR1 kinase 1b6cB. HANSWURST classifies this kinase as most closely related to 1m14A. Again, the structural superposition explains the result. The two structures are very similar. Additionally, HANSWURST clusters 1kwpA and 1csnA in one branch. Both kinases share most of the conserved structural features. Nevertheless, the structural superposition and the corresponding alignment of the sequences indicate that we aligned the proteins in a suboptimal manner. The p21-activated kinase 1 (PAK1) 1f3mC is placed close to the cluster of 1omwA and 1muoA. The second "other" kinase 1o6yA appears to be clustered with CAMK group kinases. A detailed analysis of alignments is given by Lenz[8].

To summarise, the tree is more than reasonable without any serious misalignments. Compared to the previous literature attempt[5] at classification, there are some differences, especially with the AGC family, but they appear justified. We should only agree with the literature classification of the CMGC and AK groups. Furthermore, our methodology handled an order of magnitude more structures and was fully automatic.

4 Conclusion

This work concentrated on kinases, since one can compare against literature classifications and biochemical data. Because it is fully automatic and scales well, it can be expanded to even larger families. This means we are now considering even more distantly related proteins and testing the approach on other large protein families.

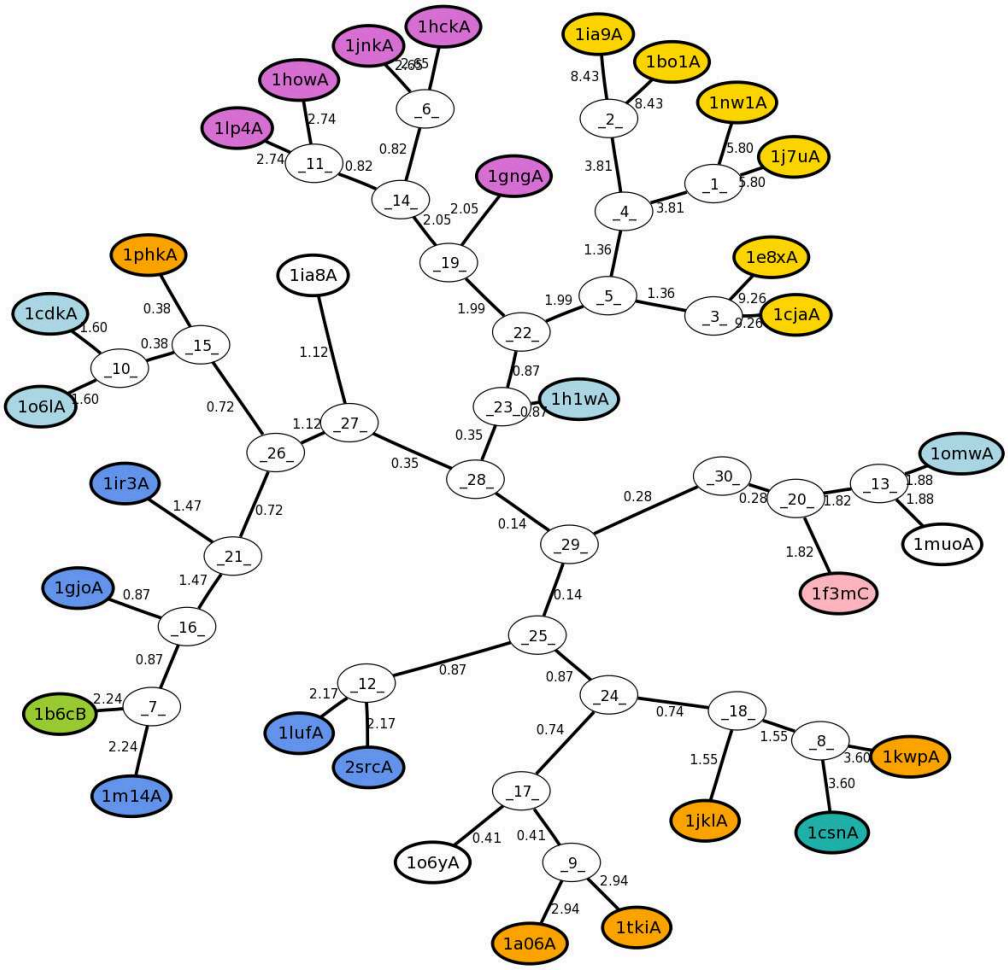


Figure 1. *HANSWURST proposed phylogeny for the literature list of recognized kinases; edges are labeled with estimated RMSD values according to clustering method; edges are not drawn to scale; atypical kinases (gold), typical kinases (cornflowerblue), AGC kinases(lightblue), tyrosine kinase like kinases (yellowgreen) depicting the most diverse group, CK1 kinases (seagreen), STE (lightpink), CMGC (magenta), CAMK (orange), uncoloured leaves depict members of the "other" group.*

References

1. S Hanks, *Genomic analysis of the eukaryotic protein kinase superfamily: a perspective.*, Genome Biol, **4**, no. 5, 111, 2003.
2. P Kennelly, *Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry.*, Biochem J, **370**, no. Pt 2, 373–389, 2003.
3. J Goldberg, G Manning, A Liu, P Fey, K Pilcher, Y Xu, and J Smith, *The dictyostelium kinome–analysis of the protein kinases from a simple model organism.*, PLoS Genet, **2**, no. 3, e38, 2006.

4. M Noble, J Endicott, and L Johnson, *Protein kinase inhibitors: insights into drug design from structure.*, Science, **303**, no. 5665, 1800–1805, 2004.
5. E Scheeff and P Bourne, *Structural evolution of the protein kinase-like superfamily.*, PLoS Comput Biol, **1**, no. 5, e49, 2005.
6. G Schenk, T Margraf, and AE Torda, *Protein sequence and structure alignments within one framework.*, Algorithms Mol Biol, **3**, no. 1, 4, 2008.
7. T Margraf, "Fast multiple alignment of protein structures", Diploma thesis, Center for Bionformatics, University of Hamburg, June 2007.
8. J Lenz, "Fast and efficient structure-based classification of kinases", Diploma thesis, Center for Bionformatics, University of Hamburg, March 2008.

Modelling of Possible Binding Modes of Caffeic Acid Derivatives to JAK3 Kinase

Jacek Kuska[1,2], **Piotr Setny**[2,3], **and Bogdan Lesyng**[2,4]

[1] Faculty of Biology, University of Warsaw,
ul. Miecznikowa 1, 02-096 Warsaw, Poland
E-mail: biochemon@bioexploratorium.pl

[2] Center of Excellence BioExploratorium, University of Warsaw,
Al. Zwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: piosto@icm.edu.pl

[3] Interdyscyplinary Center for Mathematical Modedling, University of Warsaw,
Al. Zwirki i Wigury 93, 02-089 Warsaw, Poland

[4] Faculty of Physics, University of Warsaw,
ul. Hoza 69, 00-681 Warsaw, Poland
E-mail: B.Lesyng@icm.edu.pl

Janus kinases (JAKs) belong to a family of receptor-associated protein tyrosine kinases. They play a crucial role in the JAK/STAT signaling pathways which are responsible for transduction of growth factors and cytokine-mediated signals. Abnormal activation of these pathways is observed in many types of tumors and hematopoetic malignancies[1,2] In the current study we attempt to propose binding modes of caffeic acid derivatives to JAK3, These derivatives are most likely competitive inhibitors for tyrosine-containing protein substrates. Insulin Receptor Kinase (IRK) was used to model an active, open form of JAK3. Based on JAK2 and JAK3 sequence and structure similarity analyses, residues that are responsible for the specific binding were indicated. Leading compounds were docked and modified to get a better binding specificity. The designed inhibitors are being synthesized and their biological activity will be studied experimentally.

1 Introduction

JAKs are crucial enzymes, responsible for the signal transduction through the JAK/STAT pathways, whose abnormal activation is observed in many types of tumors and hematopoetic malignancies, making them an important, but still poorly explored target for therapeutic intervention. Our goal was to create substrate competitive inhibitors of JAK3, based on our knowledge of the JAK2 inhibitors. Caffeic acid derivatives, including AG490, are the potential, promising inhibitors[3]. Even though the sequence similarity between catalytic domains of JAK2 and JAK3 is about 60%, it is difficult to propose JAK3 specific inhibitors. This is because essential residues (those which form the ATP binding pocket or the peptide binding site) are mostly the same for both kinases. In addition, there are no experimental JAK3 open (active) conformations. There is, however, an IRK structure with a bound protein inhibitor in an open enzyme conformation. Even though sequence similarity between IRK and JAKs is relatively low, the three dimensional structure similarity is significant (Fig. 1). Structural information and JAK2 and JAK3 sequence similarity analysis allowed us to identify JAK3 potential binding sites and to design specific inhibitors.

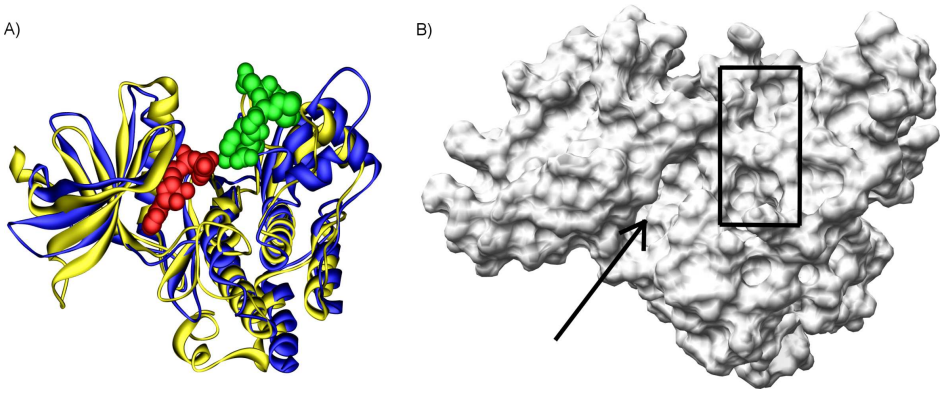


Figure 1. (A) Superposition of JAK3 on IRK. Yellow ribbon presents IRK, blue ribbon presents JAK3. The green spacefill atoms represent protein substrate atoms while the red spacefill atoms represent ANP (the ATP analog which is bound in the ATP binding cassette). (B) JAK3. The arrow indicates the nucleotide binding site, rectangle indicates the predicted docking sites.

2 Materials and Methods

Three dimensional structures of IRK, JAK3 as well as of JAK2 from the Protein Data Base were used (1IR3, 1YVJ and 2B7A structures, respectively). Because JAKs do not contain peptide substrates/inhibitors, the IRK structure in its open, active form was used for the identification of the JAK3 ligand binding site. Using MOE software we docked our potential inhibitors in the predicted peptide binding site as indicated by rectangle in Fig. 1. Because the site is relatively large, we identified two subsites that fit to our ligands. These two subsites are located in the top part and in the bottom part of the rectangle.

3 Results and Discussion

We attempted to design inhibitors that form hydrogen bonds with the backbone as well as with proteins side chains. The H-bonds with the backbone are responsible for the binding efficacy, and the H-bonds with the side chains are responsible for the substrate specificity. We indentified four potential binding sites, two of them appeared to be very promising. These are the second (Site2) and the fourth (Site4) largest pockets detected on the JAK3 surface. The most important residues which form the Site2 are: ALA952 Cys1024, ASP1025, CYS1028 and SER1031. In turn, ARG984, ASN1002 and LEU1047 form the Site4. The first binding mode accounts for interactions with the backbone of SER1029 and Ser1031. In JAK2 these positions are occupied by SER1056 and PRO1058 respectively. The second binding mode in JAK3 accounts for interactions with the backbone of ASN1002, which in JAK2 is occupied by SER1029 and SER 998. These two binding modes (Fig. 2) appeared to be the most promising. Concluding, it was possible to identify those JAK3 residues that are potentially responsible for the substrate/ligand specificity.

In order to get sufficiently high specificity of the potential inhibitors one can generate and visualize sequence variability profiles for a given protein family. Regarding protein

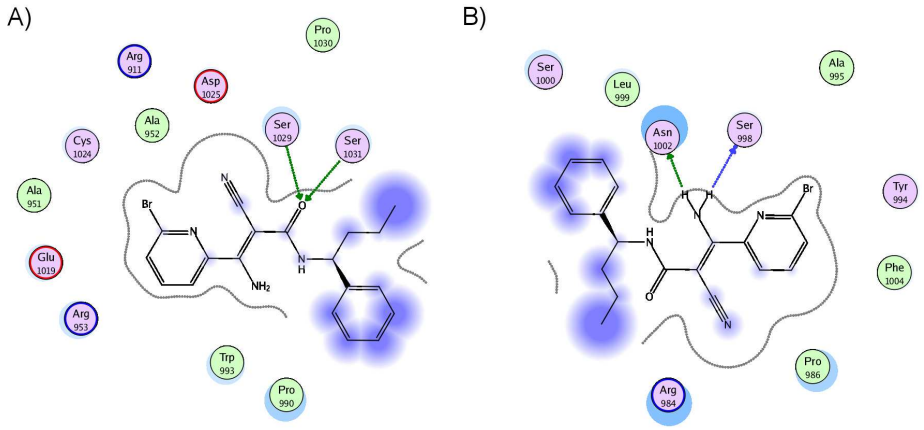


Figure 2. Binding of the caffeic acid derivatives to two subsites (A and B) in JAK3. The arrows represent the modelled hydrogen bonds that link JAK3 with added/modified atoms of the ligand.

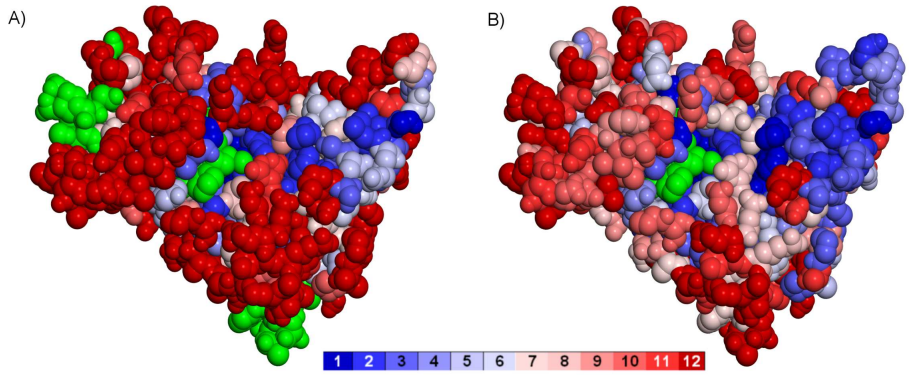


Figure 3. (A) Variability profiles resulting from alignment of 152 protein sequences from the PFAM database for the JAK family, using the JAK3 structure for visualization. (B) Visualization of 50 protein Homo sapiens sequences homologous to JAK3 acquired from the SWISS-PROT/TREMBL database. Each color represents a number of different amino acids which occupy a particular position in the multiple sequence alignment. In addition, the green color represents ATP and the protein C and N ends.

kinases, sequences from PFAM[4] and from Swiss-Prot/Trembl[5] data bases were used. A new application[6] processes these sequences and visualizes variability profiles in 3D on a 2D surface of the studied enzyme here JAK3, presented in Fig. 3. The mentioned above analysis can account for hundreds of protein sequences, it is very fast and brings a lot of essential data.

Acknowledgments

These studies were supported by the PBZ-MIN-014/P05/2005 grant as well as by Centre of Excellence BioExploratorium.

References

1. 1.H. Yu, R. Jove, *The Stats of Cancer – New Molecular Targets Come of Age*, Nature Reviews **4**, 97–106, 2004.
2. L. Gu, H. Zhuang, B. Safina, X. Xiao, W. W. Bradford, B. E. Rich, *Combinatorial Approach to Identification of Tyrphostin Inhibitors of Cytokine Signaling*, Bioorg. Med. Chem. **13**, 4269–4278, 2005.
3. P. Setny, B. Lesyng, W. Priebe, *Modelling of Possible Binding Modes of Caffeic Acid Derivatives to JAK2 Kinase* Acta Biochim. Polon. **54, Suppl. 3**, 64–65, 2007.
4. R. D. Finn, J. Mistry, B. Schuster-Bckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. L. Sonnhammer, A. Bateman, *Pfam: clans, web tools and services* Nucleic Acids Research Database **34**, D247–D251, 2006.
5. B. Boeckmann, M.C. Blatter, L. Famiglietti, U. Hinz, L. Lane, B. Roechert, A. Bairoch, *Protein variety and functional diversity: Swiss-Prot annotation in its biological context.* Comptes Rendus Biologies **328**, 882–99, 2005.
6. J. Kuska, J. Leluk, *Biow@re: a package of applications for intra/intermolecular interaction studies* Acta Biochim. Polon. **54, Suppl. 3**, 61–62, 2007.

Looking for Inhibitors of RIO Kinases

Maciej Geller[1], **Łukasz Walewski**[1,2], **Maciej Długosz**[2], **and Bogdan Lesyng**[1]

[1] CoE Bioexploratorium & Department of Biophysics, Faculty of Physics, University of Warsaw,
Żwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: mgeller@uw.edu.pl

[2] ICM, University of Warsaw,
Żwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: {ljw, mdlugosz,lesyng}@icm.edu.pl

RIO kinases are atypical protein kinases involved in ribosome synthesis. The structure of RIO2 kinase was optimized. In particular, it was virtually titrated using a Poisson-Boltzmann model. An optimal protonation state was determined at pH 7 and pH 5. Screening of a ligand database against the target protein using a docking procedure was carried out. Small flexible deformations of a binding pocket were applied. Two possible inhibitors with the best scoring function are presented. The designed leading ligands and a number of their derivatives are being synthesized and will be studied experimentally.

1 Introduction

The RIO family (RIO1, RIO2, and RIO3) of atypical serine protein kinases is conserved among archaea and eukaryotes. At least two of them, RIO1 and RIO2 are present in these organisms[1,2]. Their involvement in ribosome synthesis, a process fundamental to cell growth and proliferation, makes them attractive targets for the development of inhibitors.

There are at least three subfamilies, RIO1, RIO2, and RIO3, and it was shown that the structural features of ATP binding pockets as well as the mode of substrate binding distinguish RIO1 and RIO2 kinases. Hence, given that there is only one copy of each RIO subfamily member per organism, this should allow to design inhibitors with high specificity, which can selectively target signalling/metabolic pathways RIO kinases are involved in.

2 Methods

Crystal structure of the RIO2 (1ZAO, PDB) kinase complexed with ATP, was used. All ligands (ATP, Mn and PO4 ions, and EDO (ethylen glikol) were removed. Missing fragments of loops residues (128-130 and 135-142), as well as missing side chains of the residues (26,41, 102,131, and 133) were modeled using MM and MD methods.

There are about 108 titratable sites in the protein (15 ASP, 27 GLU, 13 TYR, 8 IS, 21 LYS, 2 CYS, 20 ARG + termini). Protonation states of the residues were determined using a well established protocol which combines:

- the Poisson-Boltzmann model for a solute-solvent system with Monte Carlo calculations[3];
- the MEAD suite[4] is used to construct the electrostatic free energy matrices, describing interactions between protein's titratable residues in their charged states;

Titratable site	Occupancy	
	pH 7	pH 5
ASP 27	0.02	0.59
ASP 196	0.01	0.46
ASP 277	0.23	0.93
GLU 112	0.03	0.53
GLU 116	0.40	0.97
GLU 180	0.03	0.56
GLU 186	0.06	0.59
GLU 244	0.06	0.73
GLU 251	0.08	0.61

Table 1.

- the DOPS program[5] uses those electrostatic free-energy matrices to compute average protonation fractions at a given pH and to generate a predefined number of protein protonation patterns with the lowest energies as found by a Monte Carlo procedure.

Dielectric constants of the solvent and proteins were set to 80 and 4 respectively. Probabilities of protonation states were evaluated at ionic strengths corresponding to 150mM of monovalent salt, at pH values in the range between 5.0 and 8.0.

At pH 7, the protonation states are the regular ones, i.e., they correspond to the protonation states of free residues. However, the experimental condition of crystallization was an acidic one, below 5 pH. The way of protonation in such conditions changes significantly. In the former, the total charge of the protein is between plus 3-5, while in the latter, between 17-18. Significant changes of the protonation states were detected in nine residues (see Table 1) none of which, however, forms the ATP binding pocket.

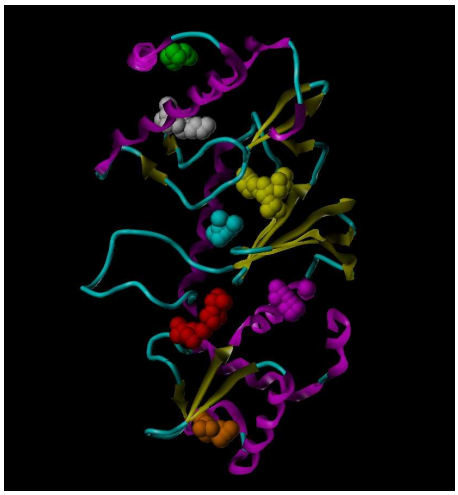
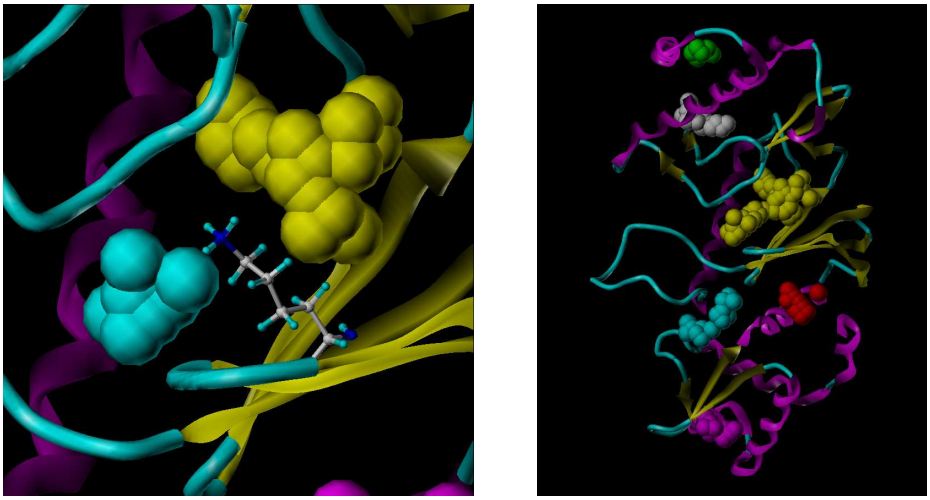


Figure 1. Location of 6 binding pockets (balls representation) of the RIO2 kinase.

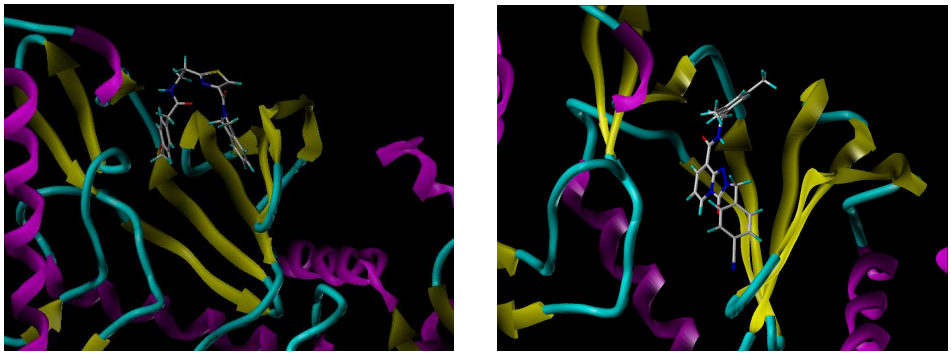


(a) (b)

Figure 2. LYS 120 side chain (a) protrudes into the binding site of the largest pocket of the RIO2 kinase (b), the yellow one.

3 Results

Possible pockets for the ligand binding are presented in Fig. 1. The ATP binding site is splitted by LYS 120 into two pockets (Fig. 2(a). These are yellow and cyan space-domains in the represetnation of the overlapping balls. Changes of the orientation of the LYS side chain results in formation of a much larger binding pocket shown in Fig. 2(b), the yellow one.



(a) (b)

Figure 3. Ligands with best score function for the RIO2 kinase.

Screening of ligands from LQ silver (`www.leadquest.com`) for the binding (Sybyl 7.3 Tripos Inc.) to the largest pocket were carried out. For comparison the non-modified two-region pocket was also scanned.

Ligand with the best score function for the non-modified position of the side chain of LYS 120 is shown in Fig. 3(a), and the best score ligand for the larger binding site is presented in Fig. 3(b).

4 Conclusions

- Novel potential inhibitors of the RIO2 kinase were designed.
- Optimization of the protonation states of ionizable side chains is of importance for the reliable modelling of RIO kinases and for the design of their inhibitors.
- Flexibility of the active site significantly changes its binding properties.

Acknowledgments

These studies are supported by PBZ-MIN-014/P05/2005 funds and by the CoE BioExploratorium.

References

1. N. LaRonde-LeBlanc and A. Wlodawer, *The RIO kinases: An atypical protein kinase family required for ribosome biogenesis and cell progression (Review)*, BBA **1754**, 14–24, 2005.
2. M. Długosz, M. Geller and L. Walewski, *Entry to modelling protonation states of ionizable groups of RIO-2 kinases* Acta. Biochim. Polon. **54, Suppl. 3**, 61–61, 2007.
3. D. Bashford and M. Karplus, Biochemistry **29**, 10219, 1990.
4. D. Bashford and K. Gerwert, J. Mol. Biol. **4**, 473, 1992.
5. J. Antosiewicz, Biophys. J. **69**, 1344, 1995.
6. D. VanDerSpoel, E. Lindhal, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701, 2005.
7. G. A. Kamiński, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474, 2001.

1.4-DHP-Lipid Forms a Tubular Micellae

Inta Liepina[1,2], **Cezary Czaplewski**[3], **Velta Ose**[4], **Reinis Danne**[1,2], **and Gunars Dubur**[1]

[1] Latvian Institute of Organic Synthesis, Riga LV1006, Latvia
E-mail: inta@osi.lv gdubur@osi.lv

[2] Center of Drug Research, Faculty of Pharmacy, University of Helsinki, Helsinki 00014, Finland
E-mail: reinis.danne@helsinki.fi

[3] Faculty of Chemistry, University of Gdańsk, 80-952 Gdańsk, Poland
E-mail: czarek@chemik.chem.univ.gda.pl

[4] Latvian Biomedical Research and Study Centre, Riga LV1067, Latvia
E-mail: velta@biomed.lu.lv

The cationic lipid 1,1.-[3,5-bis(dodecyloxycarbonyl)-4-phenyl-1,4-dihydropyridin-2,6-diyl] dimethylene bispyridinium dibromide (1,4-DHP lipid), a gene transfection agent, formed a tubular micellae during the molecular dynamics simulation with AMBER 8.0 force field. Result was confirmed with the electron microscopy showing extended, worm-like structures.

1 Introduction

Non-viral gene delivery based on self-assembling structures is an effective medical tool. The cationic lipid 1,1.-[3,5-bis(dodecyloxycarbonyl)-4-phenyl-1,4-dihydropyridin-2,6-diyl] dimethylene bispyridinium dibromide (1,4-DHP lipid, charge q=+2) (Fig. 1) is a gene transfection agent.[1,2] The electronic structure of 1,4-DHP lipid molecule was investigated by ab initio quantum mechanics, and the supramolecular structure formed by 1,4-DHP lipid molecules was investigated by means of molecular dynamics simulation.

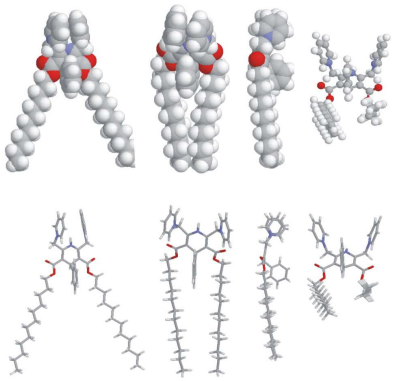


Figure 1. 1,4-DHP lipid molecule.

2 Methods, Results and Discussion

1,4-DHP lipid structure was calculated by Restricted Hartree-Fock (RHF) *ab initio* quantum mechanics, 6-31G* bases set, to obtain the charges for molecular dynamics using RESP algorithm (electrostatical potential based method using charge restrains for determining atom-centered charges). 72 molecules of 1,4-DHP-lipid were subjected to MD (AMBER 8.0 force field, NTP protocol) from the initial structure of a periodic lipid bilayer-water box, with a small amount of excessive water on the lipid edges to ensure the mobility of lipid molecules. Temperature was risen gradually from T=10 K by step of 10 degrees till 300 K. After 35 ns of MD simulation few lipid molecules turned with their charged heads to the side of the lipid bilayer and after 100 ns a profound tubular micelle structure began to form. The tubular micelle structure (Fig. 2) becomes more perfect during the course of simulation of 300 ns.

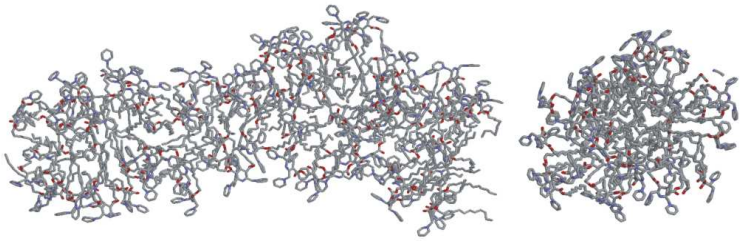


Figure 2. 1,4-DHP lipid tubomicellae side view and top view.

The results of MD simulation were confirmed by electron microscopy, showing the interwinding tubular structures (Fig. 3).

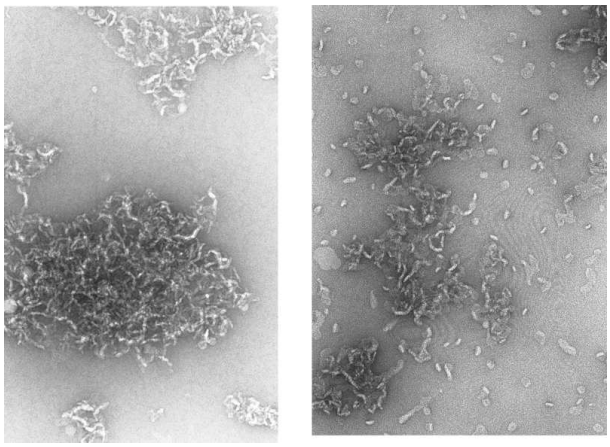


Figure 3. Electron microscopy of the 1,4-DHP lipid.

Conclusion is that one of the gene transfection agent 1,4-DHP lipid structures is a tubular micellae, and we could expect that such the micellaes are capable to form a lipoplex for the DNA transfection

Acknowledgments

This work was supported by Stipend of Finland Academy of Sciences for IL, and by Latvian Science Council Grant 05.1768. Calculations were performed on computers of the Gdansk Academic Computer Centre TASK.

References

1. Z. Hyvonen, A. Plotniece, I. Reine, B. Chekavichus, G. Duburs, A. Urtti, *Novel cationic amphiphilic 1,4-dihydropyridine derivatives for DNA delivery*, Biochim. Biophys. Acta **1509**, 451-466, 2000.
2. Z. Hyvonen, S. Ronkko, M.-R. Toppinen, I. Jaaskelainen, A. Plotniece, A .Urtti,*Dioleoyl phosphatidylethanolamine and PEG-lipid conjugates modify DNA delivery mediated by 1,4-dihydropyridine amphiphiles*, J. Controlled Release **99**, 177-190, 2004.

Protein Structure Prediction Using Coarse Grain Force Fields

Nasir Mahmood and Andrew Torda

Center for Bioinformatics, University of Hamburg,
Bundesstrasse 43, D-20146 Hamburg, Germany
E-mail: mahmood@zbh.uni-hamburg.de

In ab initio or de novo protein modelling, one tries to build 3D protein models from scratch rather than modelling them on to known structures. Our method is based on special purpose low resolution force fields. They are rather different to most approaches by not taking into account any strict physical model. They are statistical, but there is no assumption of Boltzmann statistics. In a Monte Carlo simulation, the acceptance criterion can be directly based on the calculated probabilities. Although we have not performed proper benchmark, the scoring function works reasonably well to predict 3D models of smaller proteins from their sequences.

1 Introduction

Protein structure prediction is one of the classic problems from computational chemistry or molecular structural biology. Essentially, one would like to be able to go from the sequence of a protein (easily obtained) to the structure (expensive and often difficult to obtain experimentally). Our interest has been in devising new purely probabilistic score functions. They make no use of Boltzmann statistics, but instead rely on a mixture of Bayesian probabilities based on normal and discrete distributions. This has an interesting consequence if one works with a method such as Monte Carlo, one can base the acceptance criterion directly on the calculated probabilities without assuming a Boltzmann distribution. Monte Carlo simulations, in their various forms, have been described by reputable scientists as the path to the simulator's graveyard. This poses the question as to why a rational simulator would venture further into this field. There are two aspects to this problem: 1) the score or quasi-energy function and 2) the search method. The score function may be energy-like or purely statistical and the search method is used to explore the conformational space. The score function and search method are often coupled together and search method is driven by score function to get to native like structures.

2 Method

2.1 Score Function

Unlike most Monte Carlo methods we do not use an energy or score, but calculate probabilities (or ratio of probabilities) directly:

$$Probability\,ratio = \frac{P(X_N)}{P(X_O)} \tag{1}$$

$P(X_O)$ and $P(X_N)$ are probabilities of old and updated conformations respectively.

$$P(X|\vec{V},T,S) = \prod_i [\sum_j (\pi_j \prod_k P(X_{ik}|X_i \epsilon C_j, \vec{V}_{jk}, T_{jk}, S))] \quad (2)$$

$X = X_1, ..., X_I$	the set data instances X_i (fragments)
$X_i = \vec{X_{i1}}, ..., X_{iK}$	attribute vectors X_{ik} describing X_i
i	observation number, $i = 1, ..., I$
j	class number, $j = 1, ..., J$
k	attribute number, $k = 1, ..., K$
c	inter-class probabilities and parameters
S	the set of possible probability density functions (p.d.f.) covering $\vec{V}, T$
$T = T_c, T_1, ..., T_J$	the exact functional form of each p.d.f.
$\vec{V} = \vec{V_c}, \vec{V_1}, ..., \vec{V_J}$	the set of parameters instantiating p.d.f.
π_j	class mixture probability, $\vec{V_c} = 1, ..., J$

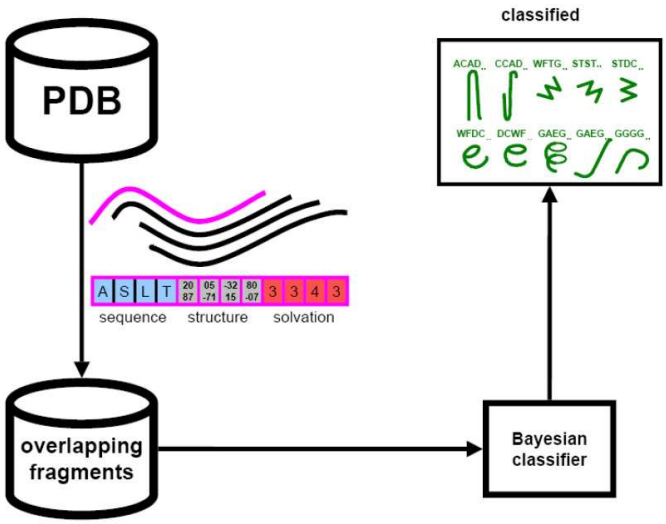


Figure 1. Bayesian classification: overlapping fragments generated from existing structures classified into a number classes by Bayesian classifier. Each fragment is represented by its sequence, structure & solvation.

Our score function is purely probabilistic and relies on mixture of Bayesian probabilities by combing sequence, structure and solvation. The statistical models: multi-way Bernoulli, bivariate Gaussian and simple Gaussian were used to model sequence, structure and solvation respectively (see figure 1).

2.2 Search Method

We are using simulated annealing Monte Carlo as a search method to find the most probable structural arrangement of a given amino acid sequence.The search method makes two kinds of moves: 1) biased moves made by drawing a fragment from a fragment library generated from existing protein structures and 2) completely unbiased moves. Internally, the score function is based on dihedral angles, Cartesian coordinates and sequence description, so there is some computational work involved in moving between representations. The acceptance criterion depends solely upon the probability ratio (equation 1) calculated from the probabilities of the new and old structures.

3 Results

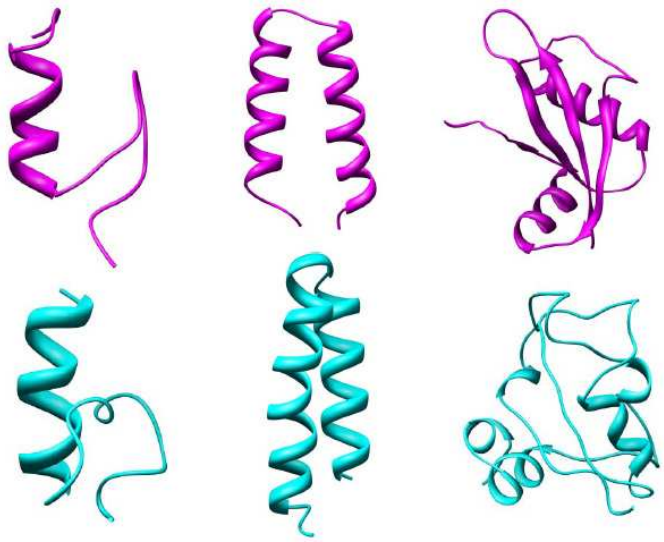


Figure 2. Top row: native structures 1fsv, 2hep and 2hfq from left to right, bottom: respective predicted models.

4 Conclusion

The current implementation seems to have a rather good representation of local interactions and works surprisingly well for small proteins. The score function has also been integrated with our existing protein threading machinery to be used for CASP8 competition. We are now working on incorporating simple solvation and hydrogen bond effects into the initial probability calculations to better account for long range interactions.

HANSWURST: Fast Efficient Multiple Protein Structure Alignments

Thomas Margraf and Andrew Torda

Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany
E-mail: {*margraf, torda*}*@zbh.uni-hamburg.de*

We have built a multiple structure alignment tool which is able to compute alignments and phylogenies of vast numbers of proteins. HANSWURST is a progressive alignment method with time complexity in the class of O(n^2). It takes advantage of a probabilistic representation of protein structure which allows for the calculation of meaningful average representations of clusters of proteins, and the alignment of those representations. Our tool scales to well over 1000 structures which is enough to cover even the largest protein families.

1 Introduction

One or two homologous sequences whisper [...]; a full multiple alignment shouts out loud.[1] This quote very eloquently describes the usefulness of multiple alignments. The significance of matches in pairwise alignments can be difficult to judge against the background noise of random matches. In multiple alignments however, random matches across a reasonable number of structures are so improbable that there is little room for doubts about their significance. This is doubly true for multiple structure alignments which really begin to shine when the relationships between proteins become so remote that sequence methods start to break down.

Common application areas for multiple structure alignments are in homology modeling[2], protein function prediction[3], creation of substitution matrices[4], phylogeny[5] and structure classification[6].

HANSWURST is built on the assumption that local interactions between atoms are the most important factor in determining the overall structure of a protein. Therefore, long stretches of high local similarity should also lead to high global similarity. From this reasoning follows that HANSWURST's aim is not to produce alignments with optimal global similarity scores such as RMSD. Instead, good global scores are considered to be a property which emerges from local similarity.

This is almost the exact opposite of the ideas behind traditional multiple structure alignment methods which sacrifice sensitivity for lower structural alignment scores.

2 Materials and Methods

This work builds on many existing methods such as AutoClass[7], rigid body superposition[8], dynamic-programming sequence alignment algorithms[9], hierarchical clustering algorithms[10], multiple sequence alignment methods and the computation of consensus probability vectors by averaging.

The basis of the alignment method is a bayesian classification of protein structure fragments using the AutoClass program[7,11].

Based on the class descriptions in such a classification, we can calculate the probability of a given protein fragment being in a certain class. The set of all class membership probabilities for a given protein fragment can be represented as a probability vector.

The dot product of two such vectors can be used as a similarity measure between two peptide fragments. This score can then be used instead of a substitution matrix in standard sequence alignment methods[9].

The resulting pairwise alignments of all vs. all structures one wishes to align are then used to fill a distance matrix. On the basis of this matrix, various clustering algorithms can be used to construct a guide tree. Currently, the best such algorithm is derived from the UPGMA method[10] and uses alignments of average probability vectors to estimate the distances between internal nodes of the guide tree. Such consensus probability vectors are computed by averaging the class membership probabilities of each fragment in a given column in the alignment. Gaps have no class memberships and thus do not contribute to the average. This concept allows each node in the guide tree to have a set of probability vectors associated with it which represent the average class memberships of that nodes descendants. Since all the information required to compute a pairwise alignment is avail-

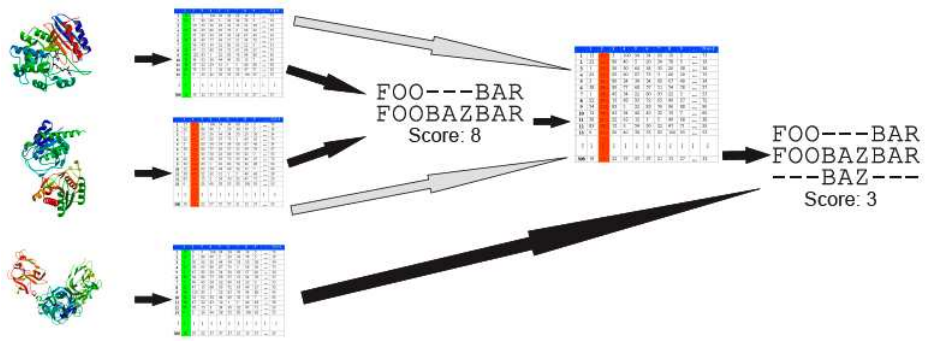


Figure 1. Illustration of the progressive construction of a three way alignment.

able for any cluster of structures, distances between clusters can be calculated by aligning their associated probability vectors. This removes the need to estimate distances during the construction of the guide tree and therefore improves its quality. The alignments of those average probability vectors are also used to merge the pairwise alignments according to the guide tree.

3 Results and Discussion

As a demonstration of our method's capabilities, we took 818 proteins with pairwise sequence identities below 25% and built a multiple structure alignment. Computing this alignment took just over 8 hours of CPU time. Anecdotal evidence suggests that this method matches almost three times as many residues as competing methods[8, 9] with some increase in the RMSD scores of the resulting superpositions. The improvement of the consensus clustering method over traditional clustering methods can be regarded as the

biggest advantage of HANSWURST over competing structural alignment methods. By representing protein structures as sets of probability vectors with regard to a fixed classification, one can easily calculate characteristic representations of clusters of proteins by averaging class membership probabilities of aligned residues in a cluster.

In combination with the method's speed and scalability, this enables the creation of multiple structure alignments of vast numbers of distantly related proteins.

References

1. T J Hubbard, A M Lesk, and A Tramontano, *Gathering them in to the fold.*, Nat Struct Biol, **3**, no. 4, 313, 1996.
2. Anthony J Russell and Andrew E Torda, *Protein sequence threading: Averaging over structures.*, Proteins, **47**, no. 4, 496–505, 2002.
3. Roman A Laskowski, James D Watson, and Janet M Thornton, *ProFunc: a server for predicting protein function from 3D structure.*, Nucleic Acids Res, **33**, no. Web Server issue, W89–93, 2005.
4. Manoj Tyagi, Venkataraman S Gowri, Narayanaswamy Srinivasan, Alexandre G de Brevern, and Bernard Offmann, *A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications.*, Proteins, **65**, no. 1, 32–39, 2006.
5. D F Feng and R F Doolittle, *Progressive sequence alignment as a prerequisite to correct phylogenetic trees.*, J Mol Evol, **25**, no. 4, 351–360, 1987.
6. L Holm and C Sander, *Dali/FSSP classification of three-dimensional protein folds.*, Nucleic Acids Res, **25**, no. 1, 231–234, 1997.
7. P. Cheeseman, J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman, *AutoClass: A Bayesian Classification System*, in: Proceedings of the Fifth International Conference on Machine Learning, pp. 54–64, Morgan Kaufmann. 1988.
8. R. Diamond, *A note on the Rotational Superposition Problem*, Acta Cryst., **A**, no. 44, 211–216, 1988.
9. SB Needleman and CD Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins*, Journal of Molecular Biology, **48**, 443–453, 1970.
10. R.R. Sokal and C.D. Michener, *A statistical method for evaluating systematic relationships*, University of Kansas Scientific Bulletin, **28**, 1958.
11. G Schenk, T Margraf, and AE Torda, *Protein sequence and structure alignments within one framework.*, Algorithms Mol Biol, **3**, no. 1, 4, 2008.

Computer Simulations of Product Dissociation from the Active Site of the Anthrax Edema Factor

Leandro Martínez, Thérèse Malliavin, Michael Nilges, and Arnaud Blondel

Unité de Bioinformatique Structurale, Institut Pasteur and URA CNRS 2185,
28, rue du Dr. Roux, 75015 Paris, France
E-mail: leandro@pasteur.fr

The Anthrax Edema Factor is an adenylyl cyclase responsible for the overproduction of cyclic mono-phosphate (cAMP) from ATP causing host cell deregulation. For optimal catalytic activity, it should efficiently release the reaction products, cAMP and Pyrophosphate. Here we study the mechanisms cAMP and PPi dissociation using Locally Enhanced Sampling and Steered Molecular Dynamics simulations. Since there is no clear consensus on the number of metal ions in the catalytic site, simulations were performed in the presence of one or two cations. The simulations suggested that the presence of the second metal ion greatly impairs product dissociation supporting the hypothesis of an optimal one-ion catalytic binding site.

1 Introduction

Anthrax bacteria produce three major toxins: Protective Antigen, Lethal Factor and Edema Factor (EF). EF enters the cell bound to the Protective Antigen. Once released in the cytoplasm it binds Calmodulin (CaM).[1] Then, a large conformational change[1] activates the adenylyl cyclase function of EF which converts ATP to cyclic-AMP (cAMP) and Pyrophosphate (PPi). To efficiently catalyse the cyclization of ATP, EF must bind ATP, stabilize the transition state (TS) and rapidly release the reaction products. TS stabilization depends on the binding of Mg^{2+} ions to the catalytic site.[2,3] The crystal structures of the EF-CaM complex bound to reaction products (1SK6[3]), contains Yb^{3+} in the active site. Two Yb^{3+} binding modes are observed: an one-ion binding mode and a two-ion binding mode, which are illustrated in Figure 1. Therefore, it is not clear whether the reaction proceeds in the presence of one or two ions. The energetic balance between TS stabilization and facilitated product release must be understood to evaluate which is the optimal active site arrangement. Here, a study combining Locally Enhanced Sampling (LES)[4] and

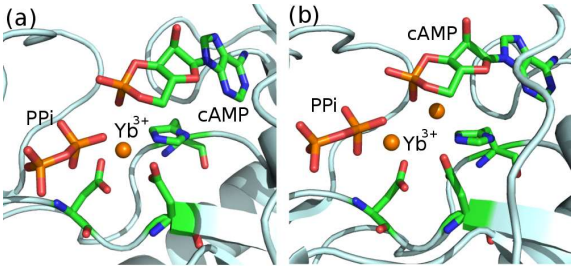


Figure 1. Ionic (Yb^{3+}) binding modes observed in the crystallographic structure of EF bound to reaction products:[3] (a) One-ion and (b) two-ion binding modes.

Steered Molecular Dynamics simulations (SMD)[5] was used to investigate the dissociation mechanisms of PPi and cAMP from the active site of the Anthrax Edema Factor under different Mg^{2+} contents. PPi and cAMP dissociated through different solvent-accessible cavities. The Mg^{2+} content of the active site greatly affected the forces required to induce product dissociation, indicating that a one-metal ion binding site would be more favorable for efficient product dissociation.

2 Molecular Dynamics Simulations of Product Dissociation

LES simulations, performed with CHARMM,[6] were used to observe dissociation of the reaction products without *a priori* assumptions on the mechanisms.[4] In this case, multiple copies of either PPi or cAMP are placed in the active site. The interaction of each copy with the protein is inversely proportional to the number of copies, and the copies do not interact with each other. We have performed 500 ps simulations with 1 to 60 copies of each of the reaction products. LES simulations were done in vacuum starting from the structure 1SK6,[3] for both ion binding modes. The ions Yb^{3+} were replaced by Mg^{2+}.

Dissociation was observed at a minimal level of 28 PPi copies. A similar number of PPi dissociation events was observed: 66 for the one-Mg^{2+} and 67 for the two-Mg^{2+} binding mode in all simulations. Dissociation of cAMP was observed with smaller number (four) of copies, and if more than 18 copies are present, all cAMP dissociated in all runs. The number of cAMP dissociation events observed for the one-Mg^{2+} binding mode was slightly larger than for the two-Mg^{2+} binding mode (77 and 61 respectively, for simulations up to 18 copies).

As shown in Figure 2, there are two discernible solvent-accessible cavities in each side of the protein Each product dissociated roughly in the directions indicated by these cavities, however displaying a significant dispersion of dissociation angles. The details of the dissociation along these directions were studied by SMD simulations.

In SMD an external force is applied to probe how difficult it is to induce ligand dissociation.[5] Simulations were performed with NAMD[7] for the fully solvated 1SK6 system, prepared with Packmol.[8] The force profiles for PPi and cAMP dissociations were (Figure 3) significantly different in in each Mg^{2+}binding mode. The presence of two ions in the active site greatly impairs product dissociation. The forces inducing product release increase from about 1500 to 3900 pN for PPi, and from 950 to 2400 pN for cAMP.

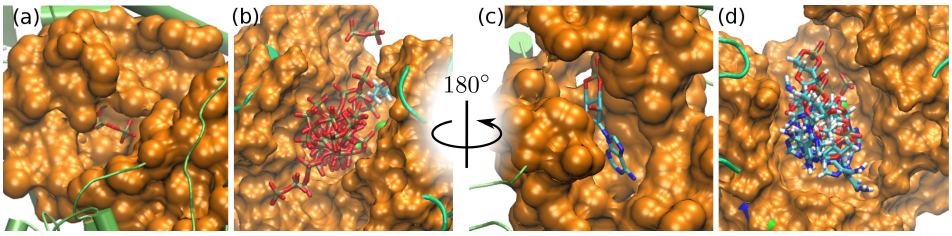


Figure 2. Product dissociation observed in LES simulations: (a) PPi solvent accessible cavity and (b) PPi dissociation. (c) cAMP solvent accessible cavity and (d) cAMP dissociation. The solvent accessible cavities are in opposite sides of the protein.

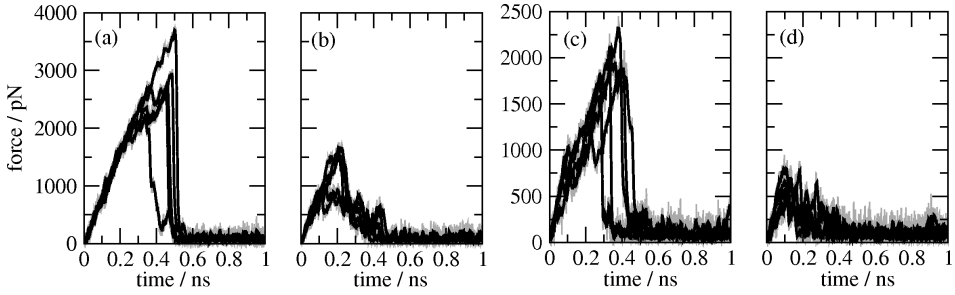


Figure 3. Forces leading to dissociation of PPi (a,b) and cAMP (c,d): five MD runs for each system where performed with variable pulling directions. The force moduli are plotted in presence of one- (a,c) and two-Mg^{2+} (b,d).

3 Concluding Remarks

The simulations suggest that the product dissociation is easier in presence of one-ion in the active site. This is in agreement with the decrease in catalytic activity observed experimentally for large Mg^{2+} concentrations[3] and suggests that the one-Mg^{2+} binding mode for EF is optimal for product release.

Acknowledgments

L. M. thanks CAPES (Brazil) for financial support.

References

1. C. L. Drum *et al. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin.* Nature **415**, 396-402, 2002.
2. Q. Shen *et al. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor.* EMBO J. **24**, 929-941, 2005.
3. Q. Guo *et al. Structural and kinetic analyses of the interaction of anthrax adenylyl cyclase toxin with reaction products cAMP and pyrophosphate.* J. Biol. Chem. **279**, 29427-29435, 2004.
4. R. Elber, M. Karplus, *Enhanced sampling in molecular dynamics: use of time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin.* J. Am. Chem. Soc. **112**, 9161-9175, 1990.
5. B. Isralewitz *et al. Steered molecular dynamics investigations of protein function.* J. Mol. Graph. **19**, 13-25, 2001.
6. B. R. Brooks *et al. CHARMM - A program for macromolecular energy, minimization and dynamics calculations.* J. Comp. Chem. **4**, 187-217, 1983.
7. J. C. Phillips *et al. Scalable molecular dynamics with NAMD.* J. Comp. Chem. **26**, 1781-1802, 2005.
8. J. M. Martínez and L. Martínez, *Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking.* J. Comp. Chem. **24**, 819-825, 2003.

A Software Library for Monte Carlo-Based Rigid Body Modelling Against Small Angle Scattering Data

Christian Meesters

Institut für Molekulare Biophysik, Universtität Mainz, 55128 Mainz, Germany
E-mail: meesters@uni-mainz.de

Rigid body modelling based on small angle scattering data is a niche in structural biology with rising importance. Here a software package is presented which tries to combine a parameter screen procedure with easy and extensible scripting capabilities. The software is intended as a *proof of concept*-software: The aim is to enable a systematic parameter screen within a reasonable time frame. This goal is reached by a built-in Monte Carlo algorithm with simulated annealing. The software allows for arbitrary symmetries & relations of protein chains. The scripting interface ensures that even "exotic cases" can be handled.

1 Introduction

In order to obtain a structural model from small angle scattering (SAS) two different approaches are followed in general: *ab initio* modelling and rigid body modelling. *Ab initio* modelling is frequently used to get a first impression of the shape of a particle, without need for *a priori* knowledge. Rigid body modelling, on the other hand, is able to provide more details, but requires initial molecular models. Several ways exist to put rigid body modelling into work with SAS data: Some require specific additional restraints, others only rely on the structural information of the model and the scattering[3]. In contrast to other programs the software presented in this article is fast, does not require restraints obtained by some other method, is not limited to particular point group symmetries, and provides an extensible scripting interface.

The theoretically most precise approach to SAS-wise rigid body modelling is computationally also the most costly one: It is possible to retrace a conformational change, by simply applying (an enormous amount of) possible moves onto the subunits. For each such conformation a check is made how well its theoretical scattering curve fits the experimental data. While this sytematic parameter screen provides an accurate estimation for the final conformation – along with uncertainties for each parameter – calculating the theoretical scattering curve many thousands of times is a severe speed bottleneck.

The presented software library divides the task into two steps: First a Monte Carlo-based search in the parameter space is performed, subsequently followed by a conventional parameter screen. In the first step a region of possibilities within the parameter space is found. The final systematic search can then be performed faster, knowing reasonable limits for the movement parameters of the bodies.

2 Software Description

The software is intentionally written as a library in the Python programming language (with some parts in C to gain speed) and not as a stand alone program. This way rapid

prototyping was granted, while potential users are enabled to easily write flexible scripts. Care was taken to have a simple & flexible user interface, while preventing the user to take nonsensical steps. The design strictly follows an object oriented paradigm throughout the construction of this library: At the heart of the library is a base class from which a class for holding atomic models is derived. This class makes use of PDB functionality provided by the biopython project[1]. In the future the base class shall provide interfaces for two classes: one holding atomic models and one, yet not released, class for holding electron densities of medium resolution (e. g. from electron microscopy or crystallographic densities of large protein complexes).

3 Modelling Strategy

The programs CRYSOL & CRYSON[5] serve as plugins for calculating theoretical scattering curves, $I(q)$, for X-rays and neutrons, respectively. In addition users can calculate the distance distribution function, $P(r)$ [see 2]. For both cases, $I(q)$ & $P(r)$, a quality factor (χ^2) can be calculated. To fit atomic 3D models to the experimental data, a grid search procedure will screen the parameter space given by rotations and translations for each independent body. The user is asked to specify geometrical restraints and movement limits in advance. The software provides options for defining bodies arbitrarily. It is possible to limit the number of atoms used for calculating the $P(r)$, while a smooth $P(r)$ is warranted. Additional scripting capabilities are provided by using free features of YASARA[4].

4 The Monte Carlo Algorithm

In order to find reasonable limits to start a systematic parameter screen (see 3) the software provides the option to perform a Monte Carlo based search beforehand (compare figure 1). Provided with SAS data an initial quality factor, χ^2, is calculated and movements for all individual chains are picked at random. To gain speed the radius of gyration (R_g) for the new model is calculated. A model is immediately rejected, if the R_g falls out of a level of tolerance. Subsequently a new χ^2 is calculated for the new model which is kept ($\chi^2_{\mathrm{new}} < \chi^2_{\mathrm{previous}}$) or rejected ($\chi^2_{\mathrm{new}} \geq \chi^2_{\mathrm{previous}}$). In order to prevent falling into a local false minimum, the algorithm might accept a "bad" move with the propability $e^{-\frac{\chi^2_{\mathrm{previous}} - \chi^2_{\mathrm{new}}}{\sqrt{\chi^2_{\mathrm{before\ cooling}}}}}$. $\chi^2_{\mathrm{before\ cooling}}$ is the discrepancy between model and experiment before "cooling down": The entire system will adjust its own movement limits and search increments, if a certain number of successful steps was reached.

5 Download & Participation

The project homepage can be found at `http://sas-rigid.origo.ethz.ch`. At the time of writing it offers access to the source code and documentation. As Python is an operating system independent language, installing Python (www.python.org) along with some freely available modules (see the documentation in a download) is sufficient to get started. – The project is an open source community project. Any kind of participation is welcome.

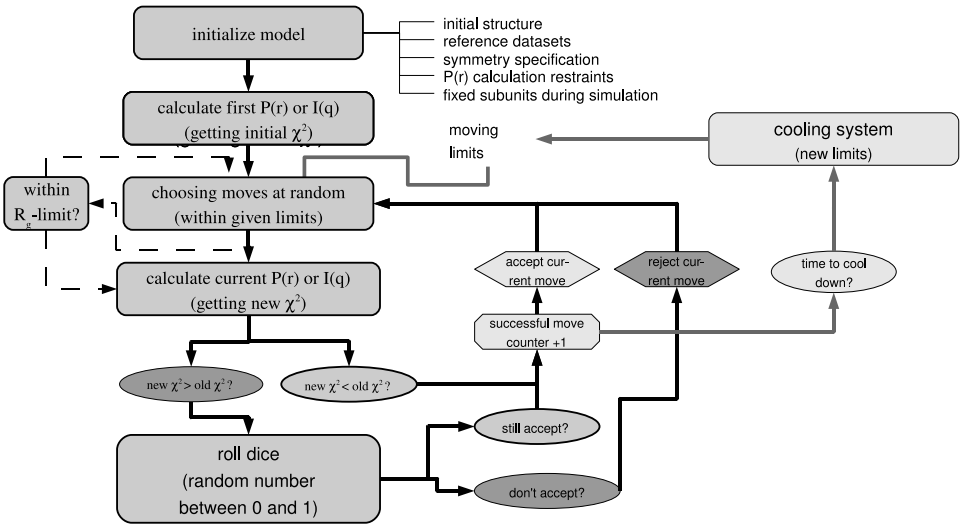


Figure 1. *Scheme of the Monte Carlo Algorithm with simulated annealing.* (See text.)

Acknowledgments

Work for this project was supported by the Sonderforschungsbereich 490. The author thanks A. Beusch, B. Pairet, and Dr. M. Binder for proofreading.

References

1. T. Hamelryck and B. Manderick, *PDB file parser and structure class implemented in Python*, Bioinformatics **19**, 2308–2310, 2003.
2. H. Hartmann and H. Decker, *All hierarchical levels are involved in conformational transitions of the 4 x 6-meric tarantula hemocyanin upon oxygenation*, Biochimica et Biophysica Acta **1601(2)**, 132–137, 2002.
3. M.B. Kozin and D. I. Svergun, *Automated matching of high- and low-resolution structural models*, Journal of Applied Crystallography **34**, 33–41, 2001.
4. E. Krieger, G. Koraimann, and G. Vriend, *Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field*, Proteins **47(3)**, 393–402, 2002.
5. D. Svergun, C. Barberato, and M. H. J. Koch, *CRYSOL - a Programm to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates*, Journal of Applied Crystallography **28**, 768–773, 1995.

Development of a Docking Methodology for Predicting the Structure of Protein-Protein Complexes

Irene Meliciani, Srinivasa Murthy Gopal, and Wolfgang Wenzel

Institute of Nanotechnology, Research Centre Karlsruhe,
76344 Eggenstein-Leopoldshafen, Germany
E-mail: {Irene.Meliciani, wenzel}@int.fzk.de

Here we report on the development of a strategy for protein-protein docking using an all-atom refinement protocol based our free-energy on-atom protein forcefield PFF02. We use a set of protein complex conformations generated by a heuristic method (ZDOCK) and score the proposed conformations using short refinement simulations in the all-atom forcefield, which predicts the experimental conformation of 1PPE (complex of bovine beta-trypsin and CMTI-I) and 1KAC (complex of knob domain from adenovirus serotype 12 and its cellular receptor CAR) to within 0.5 Å. Such an approach is much faster than a generation of the decoy set in a full simulation of the protein-complex dynamics.

1 Introduction

Many proteins are able to fulfill their biological function only in complex with other proteins. One of the most important examples for this are G-protein coupled receptors (GPCR) which bind about 40% of all known drugs. For this reason the prediction of protein-protein complexes is an important area of applications of biomolecular simulation. Free-energy based methods, such as POEM@HOME are particularly well suited to address these questions, but all-atom simulations that perform an unbiased search of all possible complex conformations are very time consuming and may not, within a reasonable time frame, even visit the experimentally relevant conformation once. One alternate approach to this problem is the generation of a decoy set, or conformational family, with a relatively inexpensive heuristic method and then to score to the members of this decoy set in the all-atom forcefield. Because models generated by one program are generally not trivially transferable to another, each of these decoys must be subjected to a short relaxation simulation. Here we investigate such a protocol, where decoys for 1PPE and 1KAC were generated with ZDOCK, which were subsequently relaxed using POEM@HOME.

2 Method

The simulations were performed with the POEM[1,2] (Protein Optimization by Energy Minimisation) simulation package using the all-atom protein forcefield PFF02[1,2], which identifies the native structure of protein/protein-complexes as the global minimum of the forcefield. The scoring approach to protein-protein docking requires two principal components (i) a fast and effective method for generating possible orientations of monomers and (ii) an accurate energy function to discriminate native and non-native conformations. We used Zdock[3] for generating the decoys set of possible conformations. A decoy set is a large library of protein conformations generated to approximately span all relevant low energy

regions of the free energy landscape. To measure the predictivity and selectivity of a force-field, the conformations in the library (decoy set) must be ranked according to their energy. If near native conformations emerge lowest in the free-energy function, the force field differentiates between native and near-native conformations. In the limit of completeness of the decoy set, which is rarely reached in practice, this test alone is sufficient to show that the force field stabilizes the native conformation of the protein against all competing metastable conformations and corresponds to the global optimum of the free-energy force field.[4] For decoy sets generated with unbiased methods, the computation of the Z-score (the difference between energies of near-native decoys to the mean energy of the decoy set in units of its standard deviation) gives a quantitative measure of the selectivity of the force field. Zdock is an initial-stage docking programme. It uses an FFT-based grid search to scan for optimal translational conformations. To complement Zdock another server Rdock[5] is used to perform the refinement of the initial-stage predictions.

3 Results

The goal of this study was to identify a suitable protocol to generate the best conformations of the complex. As a start we focused on rigid protein docking, where individual monomers do not change conformation on binding. In order to arrive at a meaningful com-

Figure 1. Docking of 1PPE to a RMSD of 0.5 Å.

parison of the energies we relaxed approximately 1000 decoys for 2 proteins in the decoy library in PFF02. This procedure maps each decoy to a local minimum of the force field of similar structure, the average change in RMSD between the starting and relaxed conformation was less than 0.02 Å, i.e. the decoys are not changed in the relaxation process.

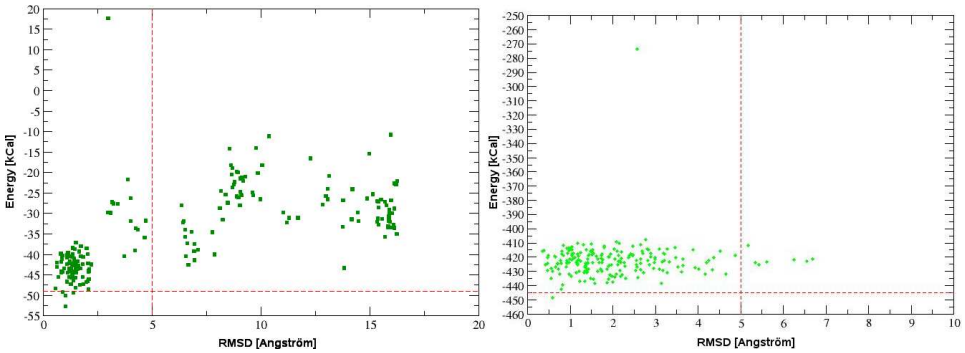


Figure 2. Results for 1PPE (left) and 1KAC (right). Free energy versus bRMSD of all accepted conformations in the simulation (25000 steps, maximum translation shift per move 0.5 Å, maximum rotational shift 0.05 rad).

Using this protocol, both proteins 1PPE (Fig. 1) and 1KAC show very good results between the native structure and the best decoys. The bRMSD-energy plot of all accepted conformations during the simulation (Fig. 2) demonstrates that the simulation explores a wide variety of conformations, with regard to their free-energy and their deviation from the native conformation. The best protocol results in 25000 steps, where we have also included rigid body translational and rotational moves to treat the aggregates. The translational moves are sampled from an equidistributed interval with maximum change of 0.5 Å, whereas the rotational moves are sampled from an equidistributed interval with maximum change of 0.05 Rad.

Acknowledgments

We thank the Leonardo da Vinci project Unipharma Graduates coordinated by Sapienza University of Rome (`www.unipharmagraduates.it`), the DFG (WE 1863/10-2) and the volunteers of POEM@HOME (`http://boinc.fzk.de`) for supporting this work.

References

1. A. Verma, W. Wenzel, *All-atom protein folding using an improved free-energy forcefield*, Biophys. J: Suppl. **S**, 214, (2007).
2. T.Herges, W. Wenzel, *An all-atom forcefield for teriary structure prediction of helical proteins*, Biophys. J **87**, 3100, (2004).
3. R. Chen, W. Tong, J. Mintseris, L. Li *ZDOCK: An Initial-stage Protein-Docking Algorithm*, Proteins **52**, 80–87, (2003).
4. C. B. Anfinsen, *Principles that govern the folding of protein chains.*, Science **181**, 223–230, 1973.
5. L. Li, R. Chen,Z. Weng, *RDOCK: Refinement of Rigid-body Protein Docking Predictions*, Proteins **53**, 693–707, (2003).

A Non-Native Helix Extension Channels Folding

Sandipan Mohanty[1] **and Ulrich H. E. Hansmann**[1,2]

[1] John v. Neumann Institute for Computing (NIC), Research Centre Jülich, 52425 Jülich, Germany
E-mail: s.mohanty@fz-juelich.de

[2] Dept. of Physics, Michigan Technological University, Houghton, MI 49931, USA

Unbiased parallel tempering Monte Carlo simulations of a 49 residue protein starting from random conformations, reveal a non-trivial path followed by the molecule to the native state. The molecule (PDB id: 2GJH) consists of an α-helix and a 3 stranded β-sheet, in which two of the adjacent strands straddle the other secondary structure elements along the sequence. In the course of folding, one of the strands making sequence non-local contacts is seen to be "cached" as a non-native extension of the native α-helix. After the other secondary structure elements have formed and assembled in their proper tertiary arrangement, the cached segment is released and it changes its secondary structure to a strand as it attaches to a β-hairpin to complete the native structure. The study is based on a physics based implicit water all-atom interaction potential called the Lund force field.

1 Introduction

While many protein structures contain β-sheets with complex arrangement of β-strands, the folding mechanisms giving rise to such structures are largely unclear. Most successful folding simulations to date have been with α-helical proteins, which are dominated by sequence local interactions. The hydrogen bonds are formed between residues i and $i+4$ along the sequence, which, even for the most stretched out conformation of the chain, are in the spatial neighbourhood of each other. Local interactions quickly lead such proteins into their folded structures. While β-sheets consisting of one or more sequence adjacent β-hairpins require somewhat longer range contacts along the sequence, they are still local structures. Typically they arise through a zipper-like mechanism starting from the turn regions and sequentially forming hydrogen bonds away from the turns. An important feature of these sequence local structures is that they do not interfere with the formation of other similar structures elsewhere along the chain.

When two neighbouring strands in a β-sheet come from regions of the sequence separated by a large number of residues, formation of contacts between them is no longer independent of the structure of the intervening segment. Premature formation of such contacts creates large steric barriers and can hinder the proper folding of other secondary structure elements. The protein is then lead into a deep local minimum and can only fold by first breaking the prematurely formed long distance native β-contacts. Such considerations suggest that proteins with complex β-sheets should on average fold slower than α-helices. This is indeed consistent with experimental observations. The so called "contact order", the average sequence separation of residues in contact, bears a striking correlation with the folding rate over a huge range of folding rates. β-sheets with complex strand arrangements have high contact orders, and are seen to fold slowly. But is it possible that some proteins with relatively high contact order might have evolved tricks to avoid the deep local minima and fold much faster than others of the same complexity? Plots of the folding rates versus

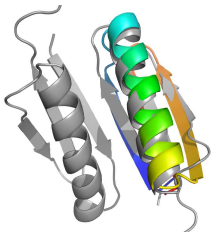


Figure 1. Comparison of the global energy minimum (colour) with the PDB structure 2GJH.pdb (grey).

contact orders suggest such a possibility, as there is a large fluctuation in the folding rates for proteins with intermediate contact orders. The exact nature of such fold accelerating mechanisms are unknown. From all-atom Monte Carlo simulations of a 49 residue protein, Top7-CFR (PDB id: 2GJH, residues 2 –50) with both helical and β-sheet structures and a non-trivial β-sheet geometry, we propose one possible mechanism.

2 Methods

Our model represents all atoms of the protein chains, including all hydrogen atoms, but an implicit treatment of the solvent molecules through an effective interaction potential. The model assumes constant bond lengths, bond angles and peptide-bond torsion angles of 180 degrees. Each protein molecule has only the Ramachandran backbone torsion angles and the side chain torsion angles as its degrees of freedom. The effective force field contains terms to account for excluded volume repulsion, local backbone electrostatics, hydrogen bonds and hydrophobic interactions. The model and the force field have been described in detail elsewhere[1,2], where we also show that the force field describes the folding and thermodynamics of a range of short peptides with both α-helical and β-sheet structures. Sampling of protein conformations is carried out using replica exchange Monte Carlo techniques with 32 replicas. For this work, we have used the protein folding software package PROFASI[3], version 1.1.2. The results presented here are based on 1.4×10^{10} elementary Monte Carlo updates of the protein chain per replica. All simulations were initialised with random values for all degrees of freedom and different random number seeds.

3 Results

The simulated molecule folds to the native state with a backbone RMSD (all residues) of about 1.8 Å. The global energy minimum found in the simulations has a backbone RMSD of 1.7 Å, and is shown superimposed on the PDB structure in Fig.1. This minimum energy structure shares all the hydrogen bonds and C_α contacts with the native state.

More interesting than the fact that the molecule folds is the observed manner of formation of β-sheet contacts between the N-terminal strand and the C-terminal hairpin[4]. When the molecule folds from random conformations, the first structures to emerge are the native helix and the C-terminal hairpin. These are strong structural elements consisting of sequence local contacts and fold to the same structures as excised segments in simulations.

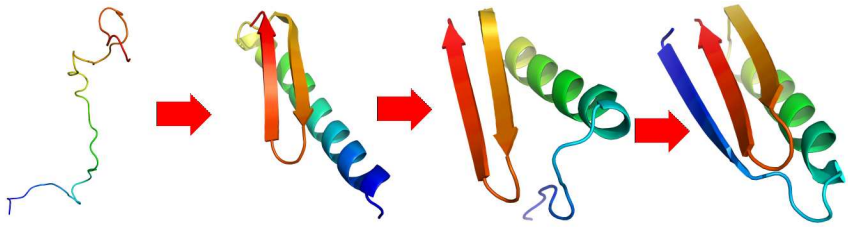


Figure 2. The folding events follow a systematic pattern, and indicate that folding is facilitated by the chameleon behaviour of the N-terminal segment.

But we observe that the N-terminal strand initially folds simply as a continuation of the native helix. β-strands are stabilised by inter-strand interactions. Therefore, initially when the C-terminal hairpin is absent, the N-terminal region does not have any stabilising interactions as a β-strand. The helix, which forms first, provides a good template for the N-terminal strands, and absorbs them. The helix, even with its non-native extension, is a structure that folds and unfolds easily. The β-hairpin forms independently, and subsequently makes hydrophobic contacts with the helix. Upon the formation of hydrophobic contacts between the helix and the hairpin, both the structures are stabilised. The non-native extension of the helix, containing the N-terminal strand residues, does not benefit from the hydrophobic contacts with the hairpin and eventually unfolds. Unlike the situation for an entirely unfolded molecule, when the N-terminal residues are freed with both the helix and hairpin in place, they do have other β-strands to bind to, which turns out to be lower in energy. Hence, the N-terminal strands join it with a larger probability.

4 Conclusions

Using all-atom Monte Carlo simulations starting from random initial conformations, we find that the molecule Top7-CFr folds to within 1.7 Å of its native state, following a non-trivial folding pathway. The observed mechanism of formation of sequence non-local β-sheet contacts depends on the chameleon behaviour of the N-terminal strand. We believe that such caching of β-strands in neighbouring helices is one mechanism for accelerating the formation of complex β-sheet structures. More detailed results are published in Refs. 4, 5.

References

1. A. Irbäck and S. Mohanty *Biophysical Journal* **88** (2005) 1560–1569.
2. S. Mohanty and U.H.E. Hansmann, *Biophysical Journal* **82** (2006) 3573.
3. A. Irbäck and S. Mohanty *Journal of Computational Chemistry* **27**, (2006) 1548. `http://cbbp.thep.lu.se/activities/profasi/`.
4. S. Mohanty, J. H. Meinke, O. Zimmermann and U.H. E. Hansmann, *Proc. Natl. Acad. Sci. U.S.A.* (2008) Published online. Doi: 10.1073/pnas.0708411105.
5. S. Mohanty and U.H.E. Hansmann, *Caching of a Chameleon Segment facilitates Folding of a Protein with End-to-End β-sheet*, submitted for publication.

Modulation of Aggregate Size and Shape Distributions of Amyloid-β Peptide Solutions by a Designed β-Sheet Breaker

Luitgard Nagel-Steger[1], Borries Demeler[2], Katrin Hochdörffer[3], Thomas Schrader[3], and Dieter Willbold[1,4]

[1] Institute for Physical Biology, Geb.26.12.U1, Heinrich-Heine-Universität Düsseldorf, 40225 Düesseldorf, Germany
E-mail: nagelst@biophys.uni-duesseldorf.de

[2] Dept. of Biochemistry, University of Texas, Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX, USA
E-mail: demeler@biochem.uthscsa.edu

[3] Institute for Organic Chemistry University Duisburg-Essen, 45117 Essen, Germany
E-mail: schrader@uni-due.de

[4] Institute for Neurosciences and Biophysics, INB-2, Research Centre Jülich, 52425 Jülich, Germany
E-mail: d.willbold@fz-juelich.de

A peptide with 42 amino acid residues (Aβ(1–42)) plays a key role in the pathogenesis of the Alzheimer's disease. It is highly prone to self aggregation leading to the formation of fibrils which are deposited in so-called amyloid plaques in the brain of affected individuals. In our study we established a method to analyze the aggregation behavior of the amyloid-β peptide with a combination of sedimentation velocity centrifugation and enhanced data evaluation software as implemented in the software package Ultrascan. Important information which becomes accessible by this methodology is the s-value distribution and concomitantly also the shape-distribution of the peptide aggregates generated in the process of self-association. These informations get especially valuable upon evaluating the properties of potential aggregation inhibitors. With this method we characterized the aggregation modifying effect of a small organic molecule, designed as a β-sheet breaker. This compound is built from three head-to-tail connected aminopyrazole moieties and represents a derivative of the already described Tripyrazole. The compound showed reduction of aggregate formation measured by FCS and decreased amyloid formation as measured by Thioflavin T measurements. By addition of this compound to a solution of the Aβ(1–42) peptide the maximum of the s-value distribution calculated for the formed amyloid-β aggregates experienced a clear shift to smaller s-values as compared to solutions where only the vehicle DMSO was added. This shift to smaller s-values was stable for at least 5 days. It could be shown that the strength of the shift was related to the amount of the added compound. The results will be discussed in terms of their significance regarding the mechanism by which the compound interferes with the fibril formation of the Aβ peptide.

1 Introduction

Protein misfolding diseases pose a major health problem not only because of their increasing incidence but especially because they still have to be regarded as incurable[1]. One of the major targets for therapy under study are the formed protein aggregates themself, whether by enhancing their clearance or the inhibition of their formation. In the case of Alzheimer's disease the misfolded component is a protein fragment generated by proteolytic cleavage of the amyloid precursor protein and consists out of 39 to 43 amino acids.

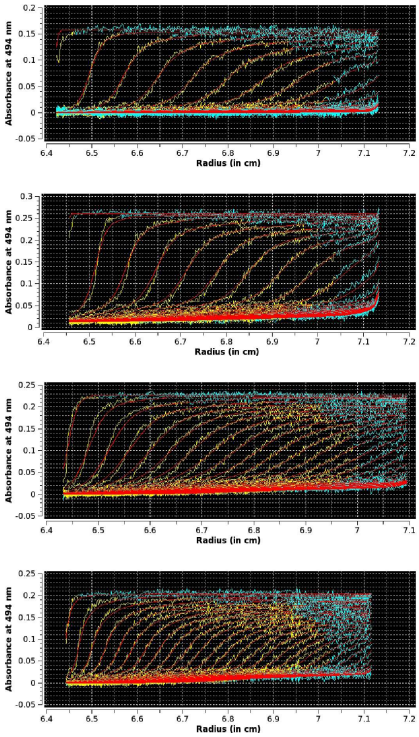


Figure 1. Raw data (yellow, blue) obtained from sedimentation velocity centrifugation of 21 μM Aβ-42/Aβ-42-OG at 20,000 rpm, 20 ℃ in 10 mM sodium-phosphate buffer, pH 7.4, 4 % DMSO. Fitted data from 2D-SA are overlaid in red. Prior to centrifugation samples were incubated slightly agitated at room temperature for 5 d. The effect of 0, 50, 150 and 200 μM compound (top to bottom) on aggregate formation is shown.

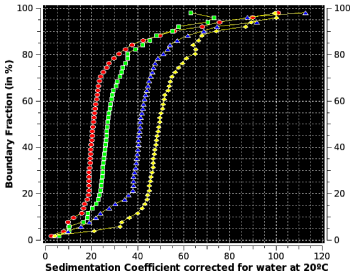


Figure 2. G(S)-distributions obtained by Van-Holde-Weischet Analysis. Yellow: 21 μM Aβ-42/Aβ-OG after 5 d incubation at room temperature without compound, with 50 μM (blue), with 150 μM (green) and with 200 μM compound (red).

The most prominent peptide is the Aβ(1–42). It is highly prone to self-association leading to different kinds of aggregates from which the mature amyloid fibril was long thought to be solely responsible for the neurodegenerative processes as observed during the course of the disease[2]. Our objectives are the development of aggregation inhibitors and the characterization of their properties *in vitro*. In previous years increasing evidence arose that probably smaller oligomeric assemblies[3,4] play a more decisive role as neurotoxic agents than the mature fibril. Information about size and shape of Aβ peptide assemblies formed during aggregation is therefore of high relevance.

Analytical ultracentrifugation is an absolute method for retrieving structural information about macromolecules by direct observation of their hydrodynamic properties in a centrifugal field. Advanced data analysis permits the determination of s-value, molecular weight and shape distributions for multicomponent systems[5]. In contrast to methods quantifying only the amyloid content of a sample, as f. e. Thioflavin T or Congo Red based fluorescence measurements, it will be possible to detect all aggregate species present in solution, from monomers to multimers consisting of several thousands units. On this account we believe that the method[6] is especially helpful in determining the effects of potential aggregation modulators. Here we present the results for a small organic compound, which is a derivative of the previously described β-sheet binder molecules consisting of aminopyrazole building blocks[7].

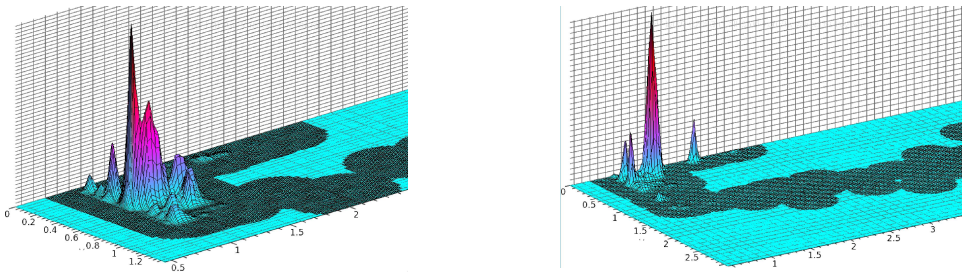


Figure 3. Results from 2-dimensional spectrum analysis of sedimentation velocity data in 3D representation. Left: 21μM Aβ 42/Aβ 42-OG in 10 mM NaP_i, pH 7.4, 4 % DMSO as control. Right: 21μM Aβ 42/Aβ 42-OG in 10 mM NaP_i, pH 7.4, 4 % DMSO with 200μM compound after 5 d incubation at RT (20,000 rpm, 20 ℃).

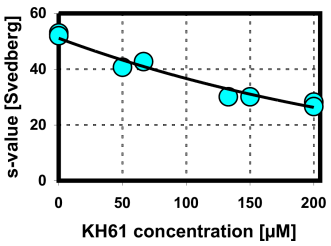


Figure 4. Dose-dependence of the inhibitory effect: Weight averaged s-values as calculated by 2D-SA/MC analyses are plotted against the compound concentration.

2 Methods

The aggregation mixture contained 17.5 μM unlabeled and 3.5 μM Oregon Green-labeled Aβ (1–42) in 10 mM sodium phosphate buffer, pH 7.4 with 4 % final DMSO concentration. DMSO was needed in order to solubilize the inhibitor compound (KH61). Samples (300 μl volume) were incubated slightly agitated at RT for 5 d prior to sedimentation velocity centrifugation. Sedimentation velocity experiments were performed with an XL-A analytical ultracentrifuge (Beckman-Coulter), equipped with absorption optics. Samples were measured in standard double-sector aluminum cells at 20,000 rpm, 20 ℃. Radial step size was set to 0.002 cm. Scans were recorded at minimal time intervals. To increase the sensitivity and the number of processable samples per run intensity instead of absorption data were recorded in continuous mode. Detection wavelength was 493 nm.

The raw data were transformed to pseudo-absorbance data, processed and evaluated using the UltraScan software package[8] running on a 44 node AMD Opteron cluster under Linux. The $\bar{v}$ value for the Aβ (1–42) as determined from the primary sequence is 0.7377 g/cm^3, the solvent density $\rho = 0.9998$ cm^3/g and viscosity $\eta = 1.0004$ centipoise. The 2-dimensional spectrum analysis (2D-SA) solves the inverse problem of fitting sedimentation velocity data to a linear combination of finite element solutions of the Lamm equation. Each term of the linear combination reflects a solute in the 2-dimensional space over s and f/f_0. Finally Monte-Carlo (MC) simulations were used to identify statistically significant solutes.

3 Results

A small organic compounds (<600 Da) designed as β-sheet binder was selected for further studies, which proved to be capable of lowering the amyloid content of Aβ 42 solutions

as measured by a Thioflavin T fluorescence assay (data not shown). The comparative sedimentation velocity analysis (Fig. 1) of Aβ 42 solutions incubated either with or without a 2 to 10fold molar excess (referring to the monomer concentration of Aβ) of the compound revealed a considerable shift of the determined s-values to smaller values as seen in the G(s) distributions determined by van-Holde-Weischet analysis (Fig. 2). This indicated an inhibited growth of aggregates caused by the added compound. An indirect effect of the compound by changing the solvent properties could be ruled out by control experiments with a protein of known s-value in the presence or absence of the compound. The results from 2D-SA/MC analysis (Fig. 4) showed the dependence of the weight averaged s-value of Aβ 42 aggregates on the applied compound concentration. Obviously it is not a single aggregate species which is stabilized by binding of the compound. More probably the measured relationship indicates an end capping mechanism of growing protofilaments or fibrils, leading to a reduced mean length of aggregates. Such a mechanism would also be expected by the design of the compound as a β-sheet binder. As can be seen in the 3D plots (Fig. 3) the number of different species is clearly reduced by compound addition, species with a frictional ratio of about 1.3 at s-values above 40 S are missing. Appropriate models for the aggregates together with further experimental data will be needed in order to interpret the determined shape related frictional ratios.

References

1. L.D. Estrada, J. Yowtak, and C. Soto. Protein misfolding disorders and rational design of antimisfolding agents. *Methods Mol. Biol.*, 340:277–293, 2006.
2. J.A. Hardy and G.A. Higgins. Alzheimer's disease: the amyloid cascade hypothesis. *Science*, 256:184–185, 1992.
3. W.L. Klein. Fibrils, protofibrils and Aβ-derived diffusible ligands: How Aβ causes neuron dysfunction and death in Alzheimer's disease. *Humana Press, Totowa, NJ*, pages 1–49, 2001.
4. D.M. Walsh, J.V. Klyubin, I. Fadeeva, W.K. Cullen, R. Anwyl, M.J. Wolfe, M.S. Rowan, and D.J. Selkoe. Natural secreted oligomers of amyloid β protein potently inhibit hippocampal longterm potentiation in vivo. *Nature*, 416:535–539, 2002.
5. E.H. Brookes, R.V. Boppana, and B. Demeler. Computing Large Sparse Multivariate Optimization Problems with an Application in Biophysics. In *SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, volume 6, page 81, 2006.
6. L. Nagel-Steger, B. Demeler, and D. Willbold. Aggregate Size and Shape Distributions in Amyloid-β Peptide Solutions. In U.H.E. Hansmann, J. Meinke, S. Mohanty, and O. Zimmermann, editors, *From Computational Biophysics to Systems Biology (CBSB07), Proceedings of the NIC Workshop 2007*, volume 36, pages 235–237. John von Neumann Institute for Computing, Jülich, 2007.
7. P. Rzepecki, L. Nagel-Steger, S. Feuerstein, U. Linne, O. Molt, R. Zadmard, K. Aschermann, M. Wehner, T. Schrader, and D. Riesner. Prevention of Alzheimer's associated Aβ-aggregation by rationally designed nonpeptidic β-sheet ligands. *J. Biol. Chem.*, 279(46):47497–47505, 2004.
8. B. Demeler. UltraScan. `http://www.ultrascan.uthscsa.edu`, vs 9.0.

Flexible Peptide-Protein Docking Employing PSO@Autodock

Vigneshwaran Namasivayam and Robert Günther

Institute of Biochemistry, Faculty of Biosciences, Pharmacy, and Psychology,
Universität Leipzig, Brüderstr. 34, 04103 Leipzig, Germany
E-mail: {vignesh, robguent}@uni-leipzig.de

In this paper we present the application of our recently developed docking program PSO@Autodock to screen a peptide library for the lethal factor of anthrax and to shed light on the application of the underlying scoring function for peptide-protein docking.

1 Introduction

Communication in biological systems occurs via specific molecular interactions. Thus, detailed knowledge of protein-ligand interactions might help to gain insight into fundamental events in the communication process of biological systems. Molecular docking methods have proven to be viable tools for studying the binding geometries and affinities of ligands to proteins. Current docking methods are designed for screening libraries of low molecular weight compounds. However, the majority of endogenous ligands are peptides. Though the development of methods to dock highly flexible ligands like peptides is evolving, the development of appropriate scoring functions is lagging behind.

We developed the molecular docking program PSO@Autodock[1] for fast flexible molecular docking. It is build upon AutoDock3 (AD3)[2], where the docking procedure is realized as multidimensional optimization. PSO@Autodock employs particle swarm optimization (PSO) techniques to find the optimal protein-ligand complex. In PSO, the locations, orientations and conformations of the ligands are represented as individual particles, which move through the search space similar to flocking birds. The *var*CPSO-ls algorithm of PSO@Autodock can efficiently screen high-dimensional search spaces. In this study we investigated the applicability of the underlying scoring function of AD3[2] for peptide-protein docking.

2 Data Preparation and Methods

For AD3 and PSO@Autodock, the protein-ligand complexes were prepared with AutoDockTools: Kollman charges were assigned for the proteins and Gasteiger charges for the ligand molecules. A grid box with a size of 90x90x90 points with a spacing of 0.375 Å was defining around the co-crystallized ligand. For GOLD (version 3.0, CCDC, Cambridge UK), the molecules were prepared using the MOE 2007.09 (CCG Inc., Montreal, Canada) and Amber89 charges were applied to the systems. Default parameters were used for AD3 and GOLD. For PSO@Autodock the cognitive and social parameters were set to 6.05. The dockings runs were stopped after 100,000 evaluations for the initial comparison study and after 500,000 evaluations for the peptide-protein dockings.

3 Flexible Peptide-Protein Docking Studies

We compared the performance of PSO@Autodock with the docking programs AD3 and GOLD. Thus, we screened a set of 10 protein complex structures with highly-flexible ligands (15 to 24 torsional angles). PSO@Autodock clearly outperforms the other docking programs. The average RMSD value of all the docked complexes is with 1.6 Å significantly lower than that obtained with AD3 or GOLD, which is above 3.0 Å in both cases (Table 1).

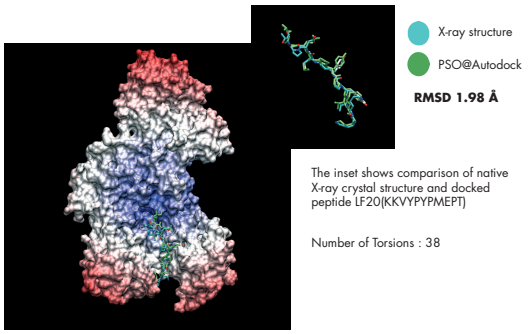
Complex	Torsions	PSO@Autodock	AutoDock3	GOLD3
1HRN	15	0.77	1.62	8.69
1CZI	16	4.39	5.71	3.03
1PPM	17	1.50	3.69	10.08
2FMB	20	3.84	4.60	6.72
3APR	21	0.66	2.87	11.54
1QRP	22	1.28	4.34	4.35
1WKR	22	0.74	1.28	11.68
3FIV	22	0.79	3.17	5.54
1HIV	23	0.85	2.98	2.77
1LYB	24	1.12	2.99	5.66

Table 1. Comparison of different docking methods (RMSD in Å).

Inspired by an experimental study[3] we applied PSO@Autodock to dock a peptide library to the anthrax lethal factor (LF)[3]. LF is a zinc-dependent metalloproteinase secreted from Bacillus anthracis that cleaves the members of the MAPK kinase (MKK) family of intracellular signaling proteins. This action by LF rapidly blocks the signals that would normally recruit other immune cells to fight the infection.

Four X-ray crystal structures of LF in complex with small molecule inhibitors and peptides have been reported[3]. First, we performed a cross-docking study in which all ligands are docked on the four crystal structures of LF to investigate whether potential ligand induced changes in the protein structure affect the accuracy of the docking. Independent of the protein structure used the native conformation of the ligand is reproduced in all complexes. The docking on the complex structure with the peptide LF20 (KKVYPYPMEPT) 1pww.pdb[3] predicts the ligands in all cases correctly with an RMSD < 2 Å (Fig. 1.a). Thus, we selected this complex for further studies.

A random sequence has been introduced in to the peptide library as a negative control as shown in Fig. 1.b. All peptides bind in the cavity region similar to the co-crystallized ligand in 1pww.pdb. Although the random sequence binds near the binding pocket, its affinity is predicted to be lower than the substrate sequences. This proves that the AD3 scoring function in PSO@Autodock can distinguish specific from non-specific peptides. However, it is difficult to differentiate between the binding strength of similar peptides. This is probably due to the limited accuracy of the AD3 scoring function[2], which has a residual error of 2.113 kcal/mol.



(a)

	Cleavage Position									
	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	ΔE^{*}_{bind}
MKK-1	K	K	K	P	T	P	I	Q	L	N	-24.16
MKK-2	R	K	P	V	L	P	A	L	T	I	-17.22
MKK-3	R	K	K	D	L	R	I	S	C	M	-21.23
MKK-4	K	R	K	A	L	K	L	N	F	A	-23.94
MKK-5	F	K	S	T	A	R	F	T	L	N	-18.72
MKK-6	R	N	P	G	L	K	I	P	K	E	-19.90
MKK-7	P	R	P	T	L	Q	L	P	L	A	-19.06
MKK-8	P	R	H	M	L	G	L	P	S	T	-19.17
Optimized	R	K	K	V	Y	P	Y	P	M	E	-20.05
Random	I	L	V	Y	G	L	S	T	V	A	-13.69

* in kcal/mol

(b)

Figure 1. (a) X-ray crystal structure of lethal factor (1pww.pdb) in complex with peptide LF20 (KKVYPYP-MEPT) (b) Peptide Library.

4 Concluding Remarks

PSO@Autodock can be applied for flexible peptide-protein docking studies. However, the scoring function currently implemented in PSO@Autodock is sufficient to discriminate between good binders and non-binders to a protein, but not accurate enough to predict the binding affinity correctly. Thus a novel scoring function has to be developed. A promising candidate for such a scoring function could follow the empirical approach of RosettaDock[4].

Acknowledgments

We are indebted to SFB 610 for financial support.

References

1. V. Namasivayam, R. Günther, *PSO@Autodock : A fast flexible molecular docking program based on swarm intelligence*, Chem. Biol. Drug. Des. **70**, 475, 2007.
2. G.M. Morris, D.S. Goodsell, R.S. Halliday, W.E. Hart, R. Belew, A.J. Olson, *Automated docking using a Lamarckian Genetic Algorithm and an empirical binding free energy function*, J. Comp. Chem. **19**, 1639, 1998.
3. B.E. Turk, T.Y. Wong, R. Schwarzenbacher, E.T. Jarrell, S.H. Leppla, R.J. Collier, R.C. Liddington,L.C. Cantley, *The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor*, Nat. Struct. Mol. Biol. **11**, 60, 2004.
4. J.J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C.A. Rohl, D. Baker, *Protein−Protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*, J. Mol. Biol. **331**, 281, 2003.

Exploration of the Energy Landscape of Protein-Protein and Antibody-Antigen Interactions

Jan Neumann[1], **Julia Morfill**[2], **and Kay E. Gottschalk**[1]

[1] Chair for Applied Physics, Ludwig-Maximilians-University,
Amalienstr. 54, 80799 Munich, Germany
E-mail: {jan.neumann, kay.gottschalk}@physik.uni-muenchen.de

[2] Center for NanoScience, Schellingstr. 4, 80799 Munich, Germany
E-mail: zimmermann@cens.de

A well choreographed, dynamic interplay of protein-protein interactions is crucial for the function of a cell. To understand these interactions, knowledge of the underlying energy landscape is essential. We analyzed the energy landscape of a protein-protein and an antibody-antigen complex using steered molecular dynamics simulations.

First, we examined the influence of velocity and geometry of the force probing on the choice of the forced unbinding pathway of the Barnase-Barstar complex. We demonstrated that in our constant velocity probing experiments, a change in the probing velocity may switch the unfolding pathway. Further more, we showed, that changing the geometry of the force probing can be used to choose between different unbinding pathways. These tools may be used for a pre-chosen sampling of the protein complex energy landscapes.

The second part of our work focused on the examination of the dependence of the unbinding pathway on the force attachment point. The truncated Leucine zipper GCN4 peptide was separated from the anti-GCN4-antibody fragment H6 in SMD simulations. Three different attachment points were examined: the C and N terminal C_αs of the 12 amino acid long peptide as well as a C_α in the middle of the peptide. We identified a common barrier on the unbinding pathway formed by a shared, central unbinding interaction.

Additionally, we classified the correlation between MD simulations and AFM as well as SPR measurements. We could show, that, in the examined system, the AFM probes the first barrier found in our MD simulations. Further more, our MD trajectories showed the existence of two main unbinding barriers. This supports the theory, that AFM and SPR may test different barriers. The second barrier is tested adiabatically with SPR measurements, while the inner barrier is probed via AFM, due to the forced tilting of the energy landscape inherent to force spectroscopy measurements.

1 Introduction

The multidimensional energy landscape of a biological molecular complex is an intrinsic property determining the dynamic function of the system. It can be described by barriers and energy minima. The barriers are blocking access to other regions of the energy landscape while the local minima define stable conformations of the complex. A system in thermodynamic equilibrium is preferentially in the global minimum of the energy profile. Disturbing the system enables sampling of the energy landscape[1,2,11]. According to the Bell-Evans model[3] the energy landscape can be tilted by applying forces. Therefore, the system is able to move over lower barriers to different regions of the energy profile. Here, we examined at the example of the Barnase-Barstar[4–7] complex[a] how the energy landscape

[a] One complex of the PDB structure 1BRS was used.

of a protein-protein complex can be more thoroughly explored. This is accomplished by using different force attachment points and different probing velocities[8]. To access larger regions of the energy landscape belonging to an antibody fragment complexed with its peptide antigen, dynamic force spectroscopy was performed by Steered Molecular Dynamics (SMD) simulations as well as Atomic Force Microscopy (AFM). To sample the equilibrium states, Surface Plasmon Resonance (SPR) was used[9].

2 Barnase Barstar Complex

2.1 Stability vs. Lability

In this experiment, we simulated the force application in analogy to AFM experiments, attaching the probing forces at single C_α-atoms[8,10]. Barstar was pulled at the last C_α of its C-terminal secondary structure element and Barnase was fixed at the first C_alpha of its N-terminal secondary structure element and vice versa. The velocity of force probing way changed from fast ($2\frac{m}{s}$ to slow ($0.5\frac{m}{s}$) pulling in these experiments, investigating the difference between probing the lability and stability of the system. If solely the system's lability is tested, it lacks time to relax in the changing energy landscape and follows the pathway of the lowest barrier heights. In case of probing only the stability of the system, the time for relaxations is sufficient and the pathway is determined only by the lowest energy. Due to the non-adiabatic pulling, the system lacks sufficient time to relax to the state of minimum energy. Hence, not only the well depth, but also the barrier height is determining the pathway. The influence of barrier heights against well depth increases with probing velocity.

2.2 Dependence on the Force Attachment Point

The differences in the anti parallel probing of the Barnase-Barstar complex lead to an extended exploration of the importance of the attachment point on the choice of the unbinding pathway. Here, typical COM measurements resulted in the unbinding of the complex only after Barstar's binding helix was separated from its protein core. Furthermore, we show, that a direct probing of the binding interface leads to a direct unbinding of the complex along two distinguishable pathways.

3 Antibody Antigen Interaction

To increase the understanding of antigen-antibody interactions, we explored the energy landscape of the antibody fragment H6 to its antigen GCN4-p1 Leucine zipper. Dynamic force spectroscopy experiments were conducted using AFM and SMD simulations to gain access to different velocity regimes. The equilibrium unbinding was determined via SPR. Probing different force attachment points on the peptide, we showed, that the unbinding under force depends on the direction of pulling, while adiabatic measurements only revealed one unbinding pathway.

The first SMD barrier rupture lengths agree with the potential widths measured by AFM. The second SMD barrier might resemble the barrier probed by SPR. These results

suggest that the system is dominated by a two barrier unbinding with a lower inner barrier and a higher outer barrier. AFM probes the inner barrier due to the applied forces tilting the energy landscape, while SPR adiabatically probes the higher outer barrier.

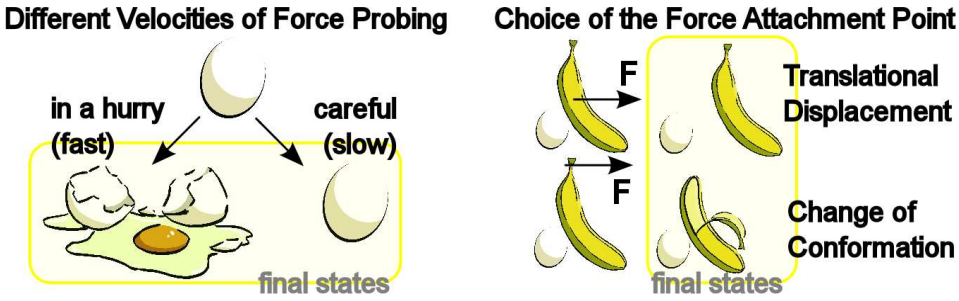


Figure 1. The exploration of the energy landscape with dynamic force spectroscopy depends not only on the choice of the attachment point of the forces, but also on the velocity of force probing. By choosing the attachment point and the probing velocity, the exploration of the energy landscape of a complex can be preferentially guided to certain regions.

4 Conclusion

The applied forces alter the energy landscape in a non-trivial way. The resulting propagation of the probed complexes through phase space does not only critically depend on the geometry of force application, but equally on the velocity of force application. Increasing the pulling velocity diminishes the time for the system to relax in the current energy landscape. The different attachment points of the force pulling alter the distribution of the tension in the complex. This results in probing different properties of the highly complex biological systems. Therefore, for a thorough characterization of the effect of force on a protein complex, multiple simulations with different probing geometries and different velocities need to be performed.

Acknowledgments

Thanks are due to Prof. R.D. Astumian and M. Höfling for a lot of helpful discussions. Special thanks to Björn Keidel for his illustrations. Further more, we thank Dr. T. Pirch and Prof. Dr. K. Jung for the SPR data. Financial support from International Doctorate Program Nano-Bio-Technology of the Elitenetzwerk Bayern, the Fonds der Chemischen Industrie, EU Project ProSurf and the Center for NanoScience is acknowledged.

References

1. V. Barsegov and D. Thirumalai, *Probing protein-protein interactions by dynamic force correlation spectroscopy.*, Phys Rev Lett, **95**, no. 16, 168302, Oct 2005.

2. R. Nevo, V. Brumfeld, R. Kapon, P. Hinterdorfer, and Z. Reich, *Direct measurement of protein energy landscape roughness.*, EMBO Rep, **6**, no. 5, 482–486, May 2005.
3. E. Evans and K. Ritchie, *Dynamic strength of molecular adhesion bonds.*, Biophys J, **72**, no. 4, 1541–1555, Apr 1997.
4. R. W. Hartley, *Barnase and barstar: two small proteins to fold and fit together.*, Trends Biochem Sci, **14**, no. 11, 450–454, Nov 1989.
5. Schreiber G. Buckle, A.M. and A.R. Fersht, *Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution.*, Biochemistry, **33**, no. 30, 8878–8889, Aug 1994.
6. G. Schreiber, A. M. Buckle, and A. R. Fersht, *Stability and function: two constraints in the evolution of barstar and other proteins.*, Structure, **2**, no. 10, 945–951, Oct 1994.
7. C. Frisch, A. R. Fersht, and G. Schreiber, *Experimental assignment of the structure of the transition state for the association of barnase and barstar.*, J Mol Biol, **308**, no. 1, 69–77, Apr 2001.
8. Hendrik Dietz and Matthias Rief, *Protein structure by mechanical triangulation.*, Proc Natl Acad Sci U S A, **103**, no. 5, 1244–1247, Jan 2006.
9. J. Morfill, J. Neumann, K. Blank, U. Steinbach, E. Puchner, KE. Gottschalk, and HE. Gaub, *Force-Based Analysis of Multidimensional Energy Landscapes: Application of Dynamic Force Spectroscopy and Steered Molecular Dynamics Simulations to an Antibody-Peptide Complex*, J Mol Biol (submitted), 2008.
10. J. Neumann, and KE. Gottschalk, *Exploring the Energy Landscape of the Barnase Barstar Complex*, Biophys J (submitted), 2008.
11. M. Kessler, KE. Gottschalk, H. Janovjak, DJ. Muller, and HE. Gaub, *Bacteriorhodopsin Folds into the Membrane against an External Force.*, J Mol Biol, **357**, no. 2, 644-654, March 2006.

From Isotope Labeling Patterns to Metabolic Flux Rates

Katharina Nöh[1], **Michael Weitzel**[2], **and Wolfgang Wiechert**[2]

[1] Institute of Biotechnology 2, Research Centre Jülich, 52425 Jülich, Germany
E-mail: k.noeh@fz-juelich.de

[2] Institute of Systems Engineering, University of Siegen, 57068 Siegen, Germany
E-mail: {michael.weitzel, wolfgang.wiechert}@uni-siegen.de

Intracellular metabolic flux rates are the manifestation of metabolic activities within organisms. Aiming at a precise quantification of the *in-vivo* fluxes, Metabolic Flux Analysis based on labeling experiments has become an valuable key technology in Systems Biology. Topology-based computational algorithms are derived to facilitate informative high-throughput analyses as well as predictive computational modeling and simulation approaches to generate new knowledge and a robust experimental design of labeling experiments.

1 Introduction

Intracellular metabolic flux rates are the most important indicators of the metabolic activities in organisms. Depending on external conditions, the *fluxome* determines the cells' physiological phenotype and, thus, their metabolic capabilities. The *in-vivo* reaction rates, however, cannot be directly accessed from measurements. Aiming at a precise quantification, Metabolic Flux Analysis (MFA) based on Isotope Labeling Experiments (ILE) became an invaluable tool for Systems Biology.

2 General Procedure: Isotope Labeling Experiments

Over the past decade two types of isotope-based MFA emerged and are being successively refined. Both methods rely on measurements of either labeled biomass components (with GC-MS, NMR) or labeled primary intermediates (LC-MS). The classical, well established *stationary* isotope MFA characterizes a cell's fluxome in a metabolic and isotopic stationary state. Typically, in a continuous culture the feed is switched from naturally labeled to isotopically labeled medium which propagates though the network and progressively replaces the unlabeled intermediates. When the labeling distribution is approximately time-invariant, samples are taken. In this field, recent experimental progress strengthened the development of high-throughput MFA[1], the investigation of extensive metabolic networks as well as the utilization of elaborated nonlinear statistical methods for flux estimation[2].

A current area of research is isotopically *non-stationary* MFA which represents a promising generalization of the classical approach[3]. Here, the cells are likewise kept under metabolic steady state conditions, however, now the time profiles of the labeling patterns are measured upon start of the labeling period in order to monitor the time-dissolved labeling propagation of the isotopic tracer though the network. Compared to classical isotope MFA, non-stationary ILEs are typically more informative and facilitate a cross check between metabolome and fluxome data. In particular for comprehensive networks, however, the required computational effort is prohibitive[3]. Altogether, the new experimental techniques result in an increased demand for more efficient algorithms.

3 Modeling and Computational Machinery

Besides the measurement of labeling patterns in key metabolites, extracellular rates, and pool sizes (non-stationary case only), a biochemical reaction network has to be provided. Additionally, atom transitions of all reactions have to be specified which essentially describe how the flow of labeled material through the cell's metabolic pathways is organized.

The analytic determination of the fluxes as a function of the emerging labeling patterns, however, is impossible for realistic networks. Thus, for unraveling intracellular fluxes, an iterative parameter fitting procedure has to be applied. Summarizing, mathematical modeling and computational algorithms build the foundation of isotope-based MFA. In this context, the computational bottleneck is the forward simulation step, i.e. the calculation of emerging labeling patterns from given parameters. Because all possible labeling combinations of each metabolite have to be considered, at the end, the solution procedure involves the solution of large, cascaded systems of Labeling Balance Equations (LBEs)[4,5].

The requirement for the computational evaluation of the model is in each case high: system dimensions range from 700 with 45 parameters (central metabolism, CM) and 5600 with 65 parameters (CM with biosynthesis pathways) up to 355.000 with 580 parameters for genome-scale models. The classical method involves the solution of an algebraic equation system while for the non-stationary method a system of ordinary differential equations, with possibly stiff characteristic, has to be solved.

4 Topological Techniques for Dimension Reduction

Efficient non-standard solution algorithms basically all rely on the structure of the Isotope Labeling Network (ILN) graph associated with the metabolic network. Although the solution algorithms for the stationary and non-stationary approaches are quite different, both approaches certainly benefit from a reduction of the problem size. Dimension reduction is performed by a careful analysis of the ILN graph followed by a removal of specified nodes and edges or a decomposition into smaller subsystems. The two basic approaches for dimension reduction are:

Path Tracing. Typically, measurement data describe only a small subset of metabolites. The forward simulation step can be restricted to a relevant sub-network which sufficiently describes the transport of labeling from the substrates to the measured metabolites[5]. The necessary path tracing procedure relies on the computation of the transitive closure of the network graphs and is performed in two directions – a *forward tracing*, which determines the fate of the isotopic labeling found in the substrates, and a *backward tracing* which determines the topological predecessors of a labeling pattern or fragment.

Network Graph Decomposition. In a divide-and-conquer approach the network graphs are decomposed into *disconnected*, unilaterally connected (*acyclic*) and *cyclic* subnetwork components (so called CCs, DAGs, and SCCs, respectively). This decomposition heavily uses the unidirectionality of reactions. Once labeling leaves a cyclically connected subnetwork (by taking an unidirectional reaction route), it is impossible for it to return. This essential information can be used for decomposition of a network into smaller subnetworks. The decomposition results in subproblems with lower dimension and, thus, dramatically reduces the running time of the solution algorithms[5]. Depending on the network connectivity the speed-up is at least in the order of two to three magnitudes.

Although the potential of these two methods is high, further reduction is possible using an even more fine-grained approach: due to their combinatorial origin, the ILNs contain isomorphic subgraphs in form of parallel paths. Since the generated LBEs are likewise isomorphic, this property results in redundant computations which are usually hard to eliminate. Fortunately, this problem is solved by the generation of analytical solutions which facilitate the elimination of common subexpressions. As a by-product analytical solutions enable exact evaluation of sensitivities and the generation of highly efficient machine code.

In case of non-stationary ILEs, typically with growing label exchange between neighboring metabolites the differential equation systems tend to be stiff. Application of e.g. a s-stage implicit Runge-Kutta scheme involves a $sN \times sN$-dimensional linear equation system, where N denotes the number of differential equations to be solved. By imposing specialized structures on the Runge-Kutta matrix, e.g. by choosing a SDIRK scheme, the computational cost for its solution can be reduced. However, the complexity of performing at least one Newton step remains. Clearly, both topological approaches presented above can be directly applied. Moreover, because the sparsitivity pattern of the Runge-Kutta matrix remains the same for all time steps, the network decomposition has to be performed only once.

5 Conclusion and Outlook

The computational routines used in stationary, and in particular non-stationary isotopic MFA suffer from the inherent computational complexity of the approach. However, exploiting the nature of the underlying (algebraic and differential) equation systems improves the efficiency of the solution methods. A careful study of the labeling network topology leads to a significant increase in performance. New algorithms emerge, having their roots in Graph Theory, Linear Algebra, and Compiler Theory. However, particularly for the classical approach the new techniques providing analytical and fast numerical solutions open the perspective to simulate even genome-scale metabolic models.

References

1. U. Sauer, *Metabolic networks in motion:* ^{13}*C-based flux analysis.* Mol. Syst. Biol. **2**, 62, 2005.
2. V. Kadirkamanathan, J. Yang, S.A. Billings, P.C. Wright PC, *Markov Chain Monte Carlo Algorithm based metabolic flux distribution analysis on Corynebacterium glutamicum.* Bioinf. **22(21)**, 2681–2687, 2006.
3. K. Nöh, K. Grönke, B. Luo, R. Takors, M. Oldiges, W. Wiechert, *Metabolic flux analysis at ultra short time scale: isotopically non-stationary* ^{13}*C labeling experiments.* J. Biotechnol. **129(2)**, 249–267, 2007.
4. K. Nöh, A. Wahl, W. Wiechert, *Computational tools for isotopically instationary* ^{13}*C labeling experiments under metabolic steady state conditions.* Metab. Eng. **8(6)**, 554–577, 2006.
5. M. Weitzel, W. Wiechert, K. Nöh, *The topology of metabolic isotope labeling networks.* BMC Bioinformatics **8**, 315, 2007.

The Inherent Stability of Collagen

Villő K. Pálfi[1] **and András Perczel**[1,2]

[1] Laboratory of Structural Chemistry and Biology,
Institute of Chemistry, Eötvös Loránd University,
Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary
E-mail: villeux@chem.elte.hu

[2] Protein Modeling Group MTA-ELTE,
Institute of Chemistry, Eötvös Loránd University,
H-1538 Budapest, P.O.B. 32, Hungary
E-mail: perczel@chem.elte.hu

Collagen is an important protein, that has an interesting secondary and tertiary structure. Three protein strands wind around each other, conserving the ε_L secondary structure. Here we explore theoretically the experimentally known thermodynamic properties of amino acid triplets on the triple helix structure. Gibbs free energy values are in complete accordance with experiments. Sarcosine (N-methyl glycine) also seems to stabilize the triple helix.

1 Introduction

Collagen is an important extracellular protein, it provides one quarter of the proteins in our body. Tropocollagen is the well-known triple helix formed by three protein chains wound around each other.[1] In each chain, glycine must be in every third place, and these places are displaced with one residue in each chain.[1] Collagen chains contain amino acid residues only in one conformation, which is characterized by the dihedral angles $\phi = -70°$, $\psi = 150°$, and is sometimes called ε_L.[1] Interestingly, the X-ray structure of polyglycine-II has the amino acids in the same conformation. According to our previous results[7], the most stable conformation for the glycine amino acids – where every amide bond is having two hydrogen bonds with two neighboring amide bonds – is this ε_L. For amino acids having an alkyl side chain, the β-pleated sheet is the most stable conformation.[7] This is the building block of amyloid, and of other plaques also, that are known to cause conformation diseases and so dementia.[3] Furthermore, it has been discovered recently that several proteins are able to form such plaques (e.g. myoglobin), after some environmental treatment.[2] Therefore plaque formation seems to be the inherent nature of proteins.[7] As a consequence, to prevent its every protein turn into a β-pleated sheet plaque, nature uses several strategies.[4] For collagen, this strategy might be the conservation of another secondary structure (the ε_L).

2 Methods

18 (3*6) residue containing collagen and triple stranded β-pleated sheet models were created as described before[8]. For all calculations the Gaussian 03[6] software was used. The structures were optimized with the "tight" criteria and subsequent frequency calculations were carried out, both at the B3LYP/6-31G(d) level of theory. Frequency calculations allow us to obtain entropy and Gibbs free energy data. The models discussed here contain

(i) only glycine amino acids (called GGG),

(ii) glycine and alanine (called AAG),

(iii) glycine and sarcosine (N-methyl glycine) (called SaSaG),

(iv) glycine and proline (called PPG),

(v) glycine and proline and hydroxyproline (called POG).

The reference values of the enthalpy and Gibbs free energy results are the sums of the enthalpies (Gibbs free energies) of the three individual peptide strands, 6 residues long each.

3 Results

Enthalpy and Gibbs free energy values of the triple helical collagen mimicking structure and the triple stranded β-pleated sheet for each amino acid composition can be seen on Table 1.

	ΔH (kcal/mol)		ΔG (kcal/mol)	
	triple helix	β-pleated sheet	triple helix	β-pleated sheet
GGG	-30.9	-68.5	27.2	-16.4
AAG	-29.0	-70.1	19.5	-29.3
SaSaG	-60.7	-44.3	-15.9	-2.9
PPG	-40.6	-9.9	0.3	21.8
POG	-40.6	-15.7	-1.3	13.1

Table 1. Enthalpies and Gibbs free energies of the triple helical and β-pleated sheet models.

For the GGG and AAG models, the sheet is the most stable structure. The Gibbs free energy of the triple helices are much higher than that of the beta sheet and also than that of the individual strands. Regarding the enthalpy values, the triple helices are more stable than the individual strands, however, the β-pleated sheet is much more stable. Experimental results[5] show that these triplets indeed destabilize the triple helix. For the SaSaG, PPG and POG triplet containing models both the enthalpy and Gibbs free energy results indicate the stability of the triple helix over the β-pleated sheet. However, a very interesting result can be seen: for the PPG and POG models the triple helix, even though it is more stable than the β-pleated sheet, has the same amount of energy as the individual strands. On the contrary, the SaSaG triple helical model is quite stable with respect to the individual strands. Unfortunately there are no experimental results for the SaSaG triplet in a triple helix, however it is shown that PPG and POG stabilizes the collagen.

4 Conclusions

In concordance with experimental results we can say that our calculations have also shown that the GGG and AAG triplets destabilize and the PPG and POG triplets stabilize the

collagen triple helix. For our great surprise sarcosine stabilizes the triple helix in a much greater extent than proline and hydroxyproline. Therefore a possible strategy for collagen to "prevent itself" from plaque formation is to eliminate the two NH amide hydrogen per three amino acids that are not included in internal hydrogen-bond formation. This can be done simply by applying a methyl group instead of the hydrogen (SaSaG). Indeed, here the triple helix is more stable than the β-pleated sheet, in a much larger extent than in the case of PPG and POG. It is very interesting, as the triple helix stabilizing effect of the proline and hydroxyproline is known for a long time. An other interesting point is that these Gibbs free energy results are in complete accordance with experimental stability data, contrary to the simple energy difference results[8].

Acknowledgments

We thank the material help of the Hungarian Scientific Research Fund (OTKA K 72973, NI-68466), MediChem2 and ICGEB (Hun04-03). The help of Svend Knak Jensen, Zoltan Szabadka and the supercomputing center of HAS KFKI AEKI and HPC Szeged is highly appreciated.

References

1. A. Bhattacharjee and M. Bansal, *Collagen structure: The Madras triple helix and the current scenario*, IUBMB Life **57**, 161, 2005.
2. C. M. Eakin, A. J. Berman and A. D. Miranker, *A native to amyloidogenic transition regulated by a backbone trigger*, Nat. Struct. Mol. Biol. **13**, 202-208, 2006.
3. J. C. C. Chan, N. A. Oyler, W. M Yau and R. Tycko, *Parallel beta-sheets and polar zippers in amyloid fibrils formed by residues 10-39 of the yeast prion protein Ure2p*, Biochemistry-Us **44**, 10669, 2005.
4. C. F. Wright, S. A. Teichmann, J. Clarke and C. M. Dobson, *The importance of sequence diversity in the aggregation and evolution of proteins*, Nature **438**, 878, 2005.
5. V. C. Chan, J. A. M. Ramshaw, A. Kirkpatrick, K. Beck and B. Brodsky, *Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix*, J. Biol. Chem. **272**, 31441, 1997.
6. M. J. Frisch et al., ,Gaussian, Inc.: Wallingford CT, 2004.
7. A. Perczel, P. Hudaky and V. K. Palfi, *Dead-End Street of Protein Folding: Thermodynamic Rationale of Amyloid Fibril Formation*, J. Am. Chem. Soc. **129**, 14959, 2007.
8. V. K. Palfi and A. Perczel, *How Stable is a Collagen Triple Helix? An Ab Initio Study on Various Collagen and β-Sheet Forming Sequences*, J. Comp. Chem. **29**, 1374, 2008.

Molecular Dynamics Simulations of the Metaloenzyme Thiocyanate Hydrolase with Non-Corrinoid Co(III) in Active Site

Lukasz Peplowski and Wieslaw Nowak

Institute of Physics, N. Copernicus University, ul. Grudziadzka 5, 87-100 Torun, Poland
E-mail: wiesiek@fizyka.umk.pl

Thiocyanate hydrolase (SCNase, pdb code 2DD5) is a novel metaloenzyme containing non-corrinoid Co^{3+} in the active site. Despite identical structure of the active sites, high sequence and structural similarity of SCNase and nitrile hydratases (NHases) both enzymes catalyse different reactions. The SCNase enzyme catalyses the degradation of thiocyanate to carbonyl sulfide and ammonia but related NHases the hydration of nitriles to amides. The main goal of the present work was to explain these different properties on the molecular level. Extensive molecular dynamics simulations (up to 10 ns) were performed using CHARMM27 forcefield with specially designed parametrization of the active site. Particular attention was devoted to water dynamics in the catalytic region, dynamical properties of the entry channel and preferential docking sites for a substrate and products of SCNase. The theoretical modelling provides useful data for understanding this enzyme having properties desired in biotechnology.

1 Introduction

Thiocyanate hydrolase (SCNase) of *Thiobacillus thioparus* THI115 is a microbial metaloenzyme with non-corrinoid CO^{3+} metal ion in the non-standard active site[1,2]. SCNase catalyses degradation of the thiocyanate to carbonyl sulfide and ammonia ($SCN^- + 2H_2O$ —> $COS + NH_3 + OH^-$)[1].

SCNase is composed of three subunits α (15 kDa), β (18 kDa) and γ (28 kDa). In γ subunit it have non-standard active site composed of four residues: γCys128, γCys131, γSer132, γCys133. The γCys131 and γCys133 residues were found to be post-translationally oxidized to cysteine-sulfinic acid $CysSO_2H$ (CSD) and cysteine-sulfenic acid CysSOH (CSO). The sequence analysis showed high similarity with quite extensively studied nitrile hydratases, especially between γ chain from SCNase and α chain from NHase. Crystallographic data show that the fold is similar to NHase and structures of the active sites are almost identical in these two enzymes[1]. Main differences in the structure of active sites is that in NHase Co^{3+} ion is six-coordinated and in SCNase it is five-coordinated. The first five ligands are identical in both cases but in NHases the sixth coordination place it is occupied by a water molecule or a hydroxide ion. In SCNase the sixth coordination place is empty[1].

Although the structure of the SCNase is known, the structure does not contain either substrate nor products of reaction. So far only in one paper[3] the interactions of the substrate and products with a mimetic center model were described. Knowledge about these interactions and dynamical properties of the SCNase can give answer about catalytical mechanism not only in thiocyanate hydrolase but also in NHase.

In this paper, for the first time, we describe docking studies of the substrate and products inside SCNase, Newtonian molecular dynamics of the enzyme and a variant of the MD

method - Locally Enhanced Samplig (LES) simulations with one of the products (NH_3) inside the protein matrix.

2 Methods

Docking was performed using the AutoDock 3.05 code[4]. Hydrogens were added into protein (chains G H I from the pdb file) using NAMD 2.6 psfgen tool[5] and minimized in NAMD 2.6. Ligands were optimised using Gaussian03 code[6] with the HF method and 6-31G* basis set. From these calculations also charges had been taken. Using AutoDockTools nonpolar hydrogens were merged and Gasteiger charges added. 256 runs of genetic algorithm (GALS) have been made for every ligand. The search grid covered the whole protein.

All molecular dynamics simulations for the enzyme were performed using NAMD 2.6 code. For the protein alone 11.5 ns long Langevin simulations in 300K with PBC have been obtained (preceded by 100ps equilibration of water box; 76 x 66 x 84 $Å^3$, and 50 ps of heating up to 300K). LES[7] simulations for the complex SCNase-ammonia were 5 ns long, with 10 copies of the ligand.

All analysis was performed using the VMD 1.8.6 code[8] and home made scripts.

3 Results and Discussion

Docking of the product (SCN^-) and substrates (COS and NH_3) show that SCN^- and COS can directly bind to the metal ion (distance Co-N 2.7 Å and Co-O 2.83 Å respectively). In these two ligands many docks exhibit also coordination to βArg90 and βArg91 located above the modified γCys131 and γCys133 (3 Å). In the case of SCN^- only nitrogen atom interacts with these arginines, but COS is stabilized by sulphur and oxygen atoms. These docks are similar to nitrile docks in NHase described in our previous paper[9]. Ammonia does not exhibit direct coordination to the metal ion. The distance in the closest docks between Co^{3+} and nitrogen is 4.65 Å. The majority of docks "near" active site are very close to γCys131, γCys133 and βArg91. There also occur docks near γArg136. We also observe docks in the channel leading to the active site, between βArg111 and βTrp112. Docks closest to the active site are shown in Figure 1.

11.5 ns simulations of the enzyme without any ligand showed that the protein is stable. Maximum value of the RMSD was 1.9 Å. Fluctuations correlate well with B factors. In this type of the MD trajectory we focused on dynamical properties of water in the neighbourhood of the active site. Analysis showed that on the distance of 5 Å from Co^{3+} ion at least 40 waters molecules can be found. The data were collected from over 260000 counts. On distance 3 Å only 10 water molecules have been found and about 4900 counts. The shortening of the distance to 2.5 Å gives 7 water molecules and only 288 counts. This indicates that the active site is accessible for water molecules, but highly hydrophilic cavity causes that water molecules occupy space in some distance from the metal ion. Such behaviour may suggest that in the catalytical mechanism SCN^- coordinates to the metal ion and in this way it is activated for a water molecule attack. This is in accordance with docking results for the SCN^-.

The purpose of the LES simulations was to find a channel leading to the cavity. Ten non-interacting copies of the ammonia were located nearby metal ion. Two of them leaved

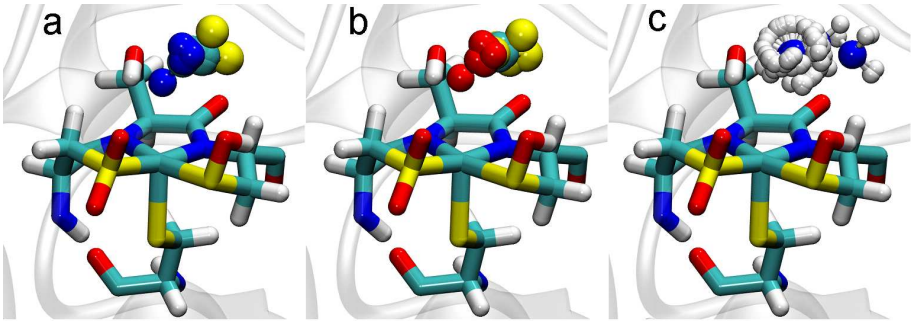


Figure 1. Docking places in the closest neighbourhood of the active site; SCN^- (a), COS (b) and ammonia (c).

the enzyme's interior. Paths of these ammonia molecules are shown in Figure 2a. The first molecule (Figure 2a black and 2b) passed to solvent after 0.7 ns. This trajectory seems to be not plausible because in the hetero-dodecameric structure this exit is closed by the other hetero-trimer and looks like "ballistic" one in the initial part of the simulation (rare artefact of the LES method). More probable is the trajectory shown in light grey (Figure 2a and 2c). In this case molecule leaved protein matrix after 1.9 ns. Contacts with residues in simulations are shown in Figure 2c. These two paths are quite different than those in NHase enzyme describes in our previous paper[10].

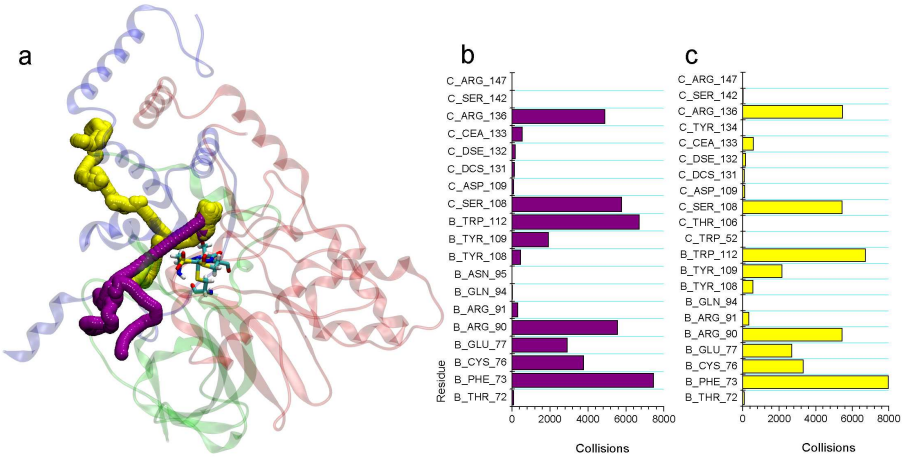


Figure 2. Trajectories of the two ammonia molecules (a), collisions counts of these molecules less plausible (b) and more plausible (c) in LES 10 simulations.

4 Conclusions

Docking studies on product and substrate show that the hypothetical catalytic mechanism relay on the binding of nitrogen atom from the SCN^- to Co^{3+} ion and such activated substrate is attacked by a water molecule. MD simulations show that a lot of water molecules are present in range of 5 Å from the metal centre. This is another hint for such type of the substrate activation. Through the LES simulation of the enzyme ammonia complex we obtained a realistic pathway of the product and indicated residues building the channel to the active site.

Acknowledgments

This research was supported by Polish Ministry of Education and Science, grant no. 2P04A 07229 and grant supported by Marszalek Wojewodztwa Kujawsko-Pomorskiego "Krok w przyszlosc - stypendia dla doktorantow". We thank Lukasz Dams for writing the script for collision counting.

References

1. T. Arakawa, *et al*., *Structure of thiocyanate hydrolase: a new nitrile hydratase family protein with a novel five-coordinate cobalt(III) center*, J. Mol. Biol. **366**, 1497, 2007.
2. Y. Katayama, *et al*., *Thiocyanate hydrolase is a cobalt-containing metalloenzyme with a cysteine-sulfinic acid ligand*, J. Am. Chem. Soc. **128**, 728, 2006.
3. J. Shearer *et al*., *A Co(III) complex in a mixed sulfur/nitrogen ligand environment: modeling the substrate- and product-bound forms of the metalloenzyme thiocyanate hydrolase*, Inorg. Chem. **39**, 4998, 2000.
4. G.M. Morris, *et al*., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*, J. Comput. Chem. **19**, 1693, 1998.
5. J.C. Phillips, *et al*., *Scalable molecular dynamics with NAMD*, J. Comp. Chem. **26**, 1781, 2005.
6. M.J. Frisch, *et al*., *Gaussian, Inc.*, Pittsburgh PA, 2003.
7. R. Elber, M. Karplus, *Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin*, J. Am. Chem. Soc. **112**, 9161, 1990.
8. W. Humphrey, A. Dalke, K. Schulten, *VMD - Visual Molecular Dynamics*, J. Molec. Graphics. **14**, 33, 1996.
9. L. Peplowski K. Kubiak, W. Nowak, *Insights into Catalytic Activity of Industrial Enzyme Co-Nitrile Hydratase. Docking Studies of Nitriles and Amides*, J. Mol. Model. **13**, 725, 2007.
10. L. Peplowski K. Kubiak, W. Nowak, *The Locally Enhanced Sampling Study of Large Ligands Diffusion inside Enzyme. Acrylonitrile and Acrylamide Journey in Nitrile Hydratase*, CBSB07 NIC Series **36**, 259, 2007.

Constraint Network Analysis: Exploiting the Link Between Protein Rigidity and Thermostability

Sebastian Radestock[1,2] **and Holger Gohlke**[1,2]

[1] Fachbereich Biowissenschaften, Molekulare Bioinformatik,
Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany

[2] Mathematisch-Naturwissenschaftliche Fakultät, Pharmazeutisches Institut,
Christian-Albrechts-Universität, Kiel, Germany
E-mail: {*radestock, gohlke*}*@pharmazie.uni-kiel.de*

Understanding the relationship between microscopic structure and macroscopic stability is important for developing strategies to improve protein stability in the reaction media used in industrial processes, *e.g.*, at high temperatures. Protein thermostability has been repeatedly linked to an enhanced structural rigidity of the folded state. Here, we used constraint network analysis for directly probing the rigidity of protein structures from mesophilic and thermophilic organisms along a thermal unfolding trajectory. The approach allowed for identifying structural features from which a destabilization of the structure originates upon thermal unfolding. These predictions showed a good agreement with experiment. The information might thus be exploited in data-driven protein engineering by pointing to residues that should be varied to obtain a protein with higher thermostability.

1 Introduction

Stable proteins are important for broadening the industrial applicability of enzymes.[1] Naturally occurring enzymes have usually not evolved to be tolerant to the presence of organic solvents, extremes of pH or high temperatures that might occur in industrial processes. The identification or the development of enzymes with higher stability will thus increase the adoption of biocatalytic syntheses in industrial production. Understanding the relationship between microscopic structure and macroscopic stability is essential for this. In this context, computational approaches that allow for identifying structural features from which a destabilization of the structure originates should provide valuable guidance.[2]

Of all potentially destabilizing factors that might occur in industrial production, temperature is the best studied.[1] As an approach to understand the determinants of thermostability, proteins from thermophilic organisms with optimum growth temperatures of more than 60°C have been investigated. These proteins show a substantially higher intrinsic thermostability than their counterparts from mesophilic organisms, while retaining the basic fold characteristics of the particular protein family.[3] Protein thermostability has been repeatedly linked to an enhanced structural rigidity of the folded state.[3]

2 Materials and Methods

Crystallographic models of 20 homologous pairs of mesophilic and thermophilic protein structures were collected from the Protein Data Bank (PDB).[4] Protein structures were modeled as constraint networks, where vertices represent atoms and edges represent covalent

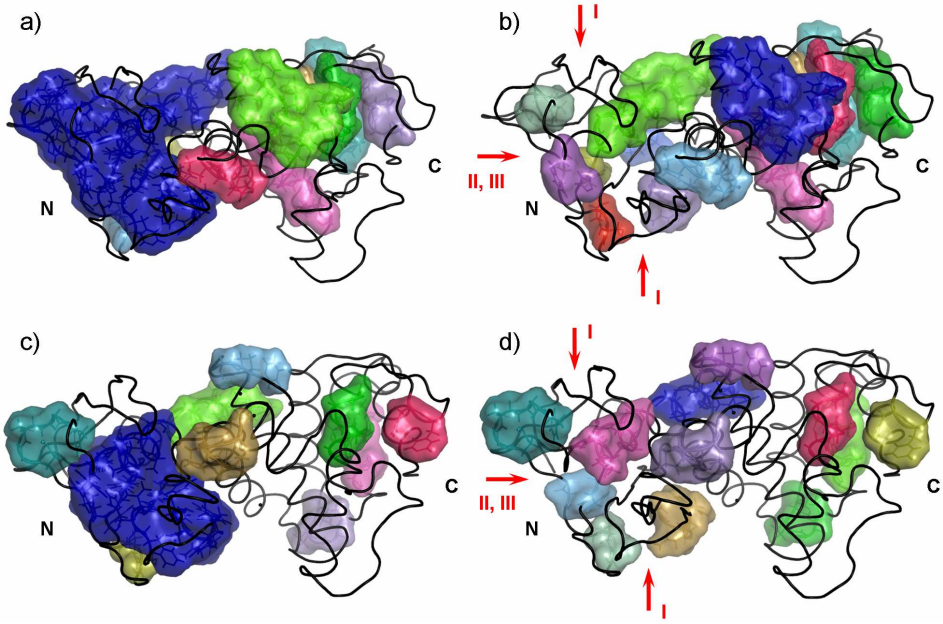


Figure 1. Mesophilic (a, b) and thermophilic (c, d) thermolysin-like protease (TLP) directly before (a, c) and after (b, d) the phase transition. Rigid clusters are depicted as uniformly colored bodies. The blue body in (a) and (c) represents the giant cluster. Arrows in (b) and (d) indicate potential unfolding nuclei. Roman numbers refer to the numbering of the unfolding nuclei in Table 1. The N- and C-termini are marked.

and non-covalent bond constraints as well as angular constraints. The network was constructued using the FIRST software (version 6.2).[5] A fast combinatorial algorithm can be applied to determine the number and spatial distribution of bond-rotational degrees of freedom in the network and, hence, the local network rigidity. Such a rigidity analysis is available with FIRST.[5]

By diluting non-covalent constraints in the protein structure network starting from the native state, FIRST has been applied to simulate thermal unfolding of proteins.[6] Here, heating was simulated by removing hydrogen bonds from the network in the order of increasing interaction energy. The energy of a hydrogen bond relates to the temperature at which the bond breaks. In going from a rigid to a flexible network, a phase transition can be observed that defines the rigidity percolation threshold. To identify the temperature of the phase transition T_p, concepts from percolation theory and network science were applied.[7,8]

3 Results and Discussion

In a first step, the general percolation behavior of the constraint networks was analyzed. The phase transition can be viewed as a rigid to flexible transition of the kind observed in network glasses. It is characterized by the decay of a large rigid cluster (the giant cluster) in the network.[4] The temperature of the phase transition relates to the melting temperature of the protein. A higher phase transition temperature was observed for approximately two-

thirds of the proteins from thermophilic organisms from our data set compared to their mesophilic counterparts (data not shown).

In a second step, the microscopic structure of the networks was related to their observed macroscopic behavior, in order to characterize stability features of the protein structures. For this, networks directly before and after the phase transition were compared. In Figure 1, the rigid cluster decomposition of mesophilic and thermophilic thermolysin-like protein (TLP) is shown. Figure 1a and c show the networks from mesophilic and thermophilic TLP directly before the phase transition, respectively. Apparently, the giant cluster dominates the system in both cases. Moreover, the giant cluster is located in the same region of the proteins: It extends over the N-terminal domain and comprises the β-sheet region and an α-helix in the N-terminal domain. After the phase transition, the giant cluster decays into smaller rigid clusters and regions that are flexible (Figure 1b and d). The close correspondence of the rigid cluster distribution in the networks of the homologous proteins before and after the phase transition is an intriguing result of our analysis.

Unfolding nucleus		**Predicted sites**	**Experimentally verified sites**[9]
I	β-sheet region in the N-terminal domain	21-24, 29, 31-34, 39-42, 44, 101-107, 114-115, 122-123	–
II	N-terminus of the α-helix in the N-terminal domain	68-70	69
III	Region around F63 in the N-terminal domain	54, 56-62	4, 56, 58, 63, 65

Table 1. Comparison of predicted with experimentally verified unfolding sites[9] for thermolysin-like protein (TLP).

In analogy to experimental protein unfolding, where initial unfolding of local regions precedes the denaturation of the entire protein, the loss of rigidity in certain regions is considered to precede the phase transition. These regions were identified as parts of the giant cluster that become flexible upon the phase transition, each representing an unfolding nucleus. In case of TLP, three unfolding nuclei could be found (Table 1, Figure 1). The predicted unfolding nuclei were compared with experimental data (Table 1). Notably, the predicted unfolding nuclei are in good agreement with sites where stabilizing mutations have successfully been introduced into TLP.[9] A likewise good agreement between our predictions and experimental data was found for many other proteins from our data set.

4 Concluding Remarks

Our findings strongly support the notion that the stability of thermophilic proteins is in general linked to an enhanced structural rigidity of the folded native state.[3] Furthermore, direct support is found for the corresponding states concept which states that homologous proteins exist in corresponding states of similar flexibility at their respective optimal temperature.[3] To the best of our knowledge, this is the first theoretical study addressing this

issue by directly probing the rigidity of protein structures along a thermal unfolding trajectory for a comprehensive dataset. Regarding the identification of regions that become flexible when approaching the phase transition (unfolding nuclei), we were encouraged to see the good (albeit not perfect) agreement between predicted sites and experimental mutations that led to higher structural stability. The result demonstrates that our approach will indeed be helpful to guide data-driven protein engineering to regions where mutations most likely will have a notable effect on thermostability.

References

1. D. C. Demirjian, F. Moris-Varas, C. S. Cassidy, *Enzymes from extremophiles*, Curr. Opin. Chem. Biol. **5**, 144–151, 2001.
2. J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *Better library design: Data-driven protein engineering*, Biotechnol. J. **2**, 180-191, 2007.
3. R. Jaenicke, G. Böhm, *The stability of proteins in extreme environments*, Curr. Opin. Struct. Biol. **8**, 738-748, 1998.
4. S. Radestock, H. Gohlke, *Exploiting the link between protein rigidity and thermostability for data-driven protein engineering*, submitted.
5. D. J. Jacobs, A. J. Rader, L. A. Kuhn, M. F. Thorpe, *Protein flexibility predictions using graph theory*, Proteins **4**, 150–165, 2001.
6. A. J. Rader, B. M. Hespenheide, L. A. Kuhn, M. F. Thorpe, *Protein unfolding: Rigidity lost*, Proc. Natl. Acad. Sci. U.S.A. **99**, 3540–3545, 2002.
7. D. Stauffer, A. Aharony, *Introduction to percolation theory*, 2nd ed., Taylor and Francis, London, 1994.
8. I. R. Tsang, I. J. Tsang, *Cluster size diversity, percolation, and complex systems*, Phys. Rev. E **60**, 2684-2698, 1999.
9. B. van den Burg, B. W. Dijkstra, G. Vriend, B. Vandervinne, G. Venema, V. G. H. Eijsink, *Protein stabilization by hydrophobic interactions at the surface*, Eur. J. Biochem. **220**, 981-985, 1994.

Solvent in Protein Interfaces: Molecular Dynamics Approach

Sergey Samsonov, Joan Teyra, M. Teresa Pisabarro

BIOTEC TU Dresden, Tatzberg 47/49, 01307, Dresden, Germany
E-mail: sergeys@biotec.tu-dresden.de

Water molecules are present ubiquitously in living cells. However, solvent contribution to protein-protein interactions is often ignored in protein-protein interactions studies. Previous work has suggested the importance of wet spots (residues interacting only through one water molecule) in description of protein interactions. We use a molecular dynamics approach to analyze solvent in protein interfaces. Our results show that residence time of water molecules in wet spots sites is found to be significantly higher than of water molecules on protein surfaces. In terms of free energy these water molecules are heterogeneous. Nevertheless, their contribution to the free energy of complex formation significantly changes the energy function of the system suggesting that water should be considered in detailed protein interface description.

1 Introduction

Water plays an extremely important role in all biological processes. Water molecules have been shown to be structurally conserved in protein complexes and to contribute thermodynamically to protein complex formation, while their residence time and diffusion characteristics in protein interfaces are distinct from bulk and surface solvent [1]. Despite all, solvent is often ignored in the analysis of protein-protein interactions.

In our previous work we have developed SCOWLP, which, taking into account interfacial solvent, classifies all interfacial protein residues of the PDB into three classes based on their interacting properties: dry (direct interaction), dual (direct and water-mediated interactions), and wet spots (residues interacting only through one water molecule) [2]. This study aims to gain insights into dynamic and energetic properties of solvent in protein interfaces using MD approach.

2 Methods

10 ns of MD productive runs were carried out in AMBER 8.0 using isothermal isobaric periodic boundary conditions and TIP3P water model. Interfacial interactions were defined by the SCOWLP criteria [2]. If the interacting heavy atoms of each wet spot counterpart were closer than 3.6 to water molecules, the wet spot site was considered to be occupied. A surface water site was defined by the volume with 100% total occupancy around one of the protein polar groups outside the interface. A bulk water site was defined similarly in terms of total occupancy. Residence time distribution density, maximum residence time (Tmax) and total occupancy were defined as the frequency of consecutively occupied frames, maximum number of consecutively occupied frames and the total time when the site was occupied, respectively. For free energy calculations, the double decoupling method of free energy perturbation was used [3]. The coupling parameter l was variated

from 0 to 1 and back with a 0.01 step. The system was was equilibrated for 10 ps for each l value followed by a 10 ps productive MD run.

3 Results

3.1 Residence Time of Water Molecules in Wet Spot Sites

The analysis of wet spot sites from 17 protein-peptide and protein-protein complexes suggests that residence time density distribution for each site is described as: $F(t) = Ct^{-k}$, where C is a normalization constant, and $k > 0$ is the only distribution parameter. k and Tmax were compared for wet spots, surface and bulk water sites. It was shown that both parameters significantly differ (at the level of t-Test p-value=0.05) for different sites, indicating that water molecules in wet spot sites are more stable than in bulk solvent or in surface hydration sites (Figure 1). At the same time, in each wet spot site there are many occupancy events that have as short residence as in bulk or surface sites. That agrees with the model proposed by Makarov et al. Here, the correlation function for residence time in hydration sites is decomposed into the sum of fast and slow diffusion exponent components, which characterize bulk water motions and specific for hydration site events, respectively [10]. Other theoretical and experimental studies obtained similar residence time values for different water sites, which vary from 1-10 ps for bulk solvent to $10^2 - 10^3$ ps for protein hydration sites, cavities and cores [1]. Tmax and k are well correlated (adjusted correlation coefficient r=0.81 for ln(Tmax) k, while there was no correlation between total occupancy of the sites and Tmax ($r < 0.3$) because these parameters are independent and describe different kinetic characteristics of the site. While Tmax is defined only by the energy barrier required for the molecule to leave the site, total occupancy is also dependent on the energy barrier of water transfer from bulk solvent to the site. The residence time analysis suggests that the potential barriers for wet spots sites are significantly higher than those for surface sites.

3.2 Free Energy of Water Molecules in Wet Spot Sites

To determine if water molecules contribute energetically favorably to complex formation we calculated their free energy. As a first step, free energy of removing a water molecule from bulk solvent was calculated. Electrostatic and van der Waals components were equal to 8.2 and -2.2 kcal/mol, respectively, which agrees well with the results obtained from similar calculations correlated with experimental data [3]. The second step consisted of the transfer of a water molecule from the wet spot site to vacuum. The difference of these two energy components makes up the total energy of a water molecule transfer from bulk solvent to the wet spot site. The obtained results for several water sites of the SH3 domain complex 1uj0 show that the sites are very heterogeneous. In particular, the free energy of water molecule transfer from the site formed by the carboxyl oxygen of Glu12 in the SH3 domain and the side-chain of Arg64 in the ligand is -1.4 kcal/mol, meaning a favorable impact of a water molecule on the complex formation. The calculations carried out for another site formed by the side-chain of Asp34 in the SH3 domain and the side-chain of Asn66 in the ligand revealed a positive change of free energy (1.3 kcal/mol). However, as it was observed in the trajectory, an additional water molecule was present in

this site and establishing a hydrogen bond with the first water molecule forming the wet spot. Consideration of both water molecules in the free energy calculations revealed an energy gain of -6.6 kcal/mol, reflecting a water cooperativity effect. Another example of cooperativity effect was found in the site formed by the side-chain of Asn52 in the SH3 domain and the main-chain of Met61 in the ligand. Here, although the energy becomes more favorable by consideration of two water molecules transfer, water contribution was still not favorable. In surface sites no big negative values for free energy were found. In other calculations using the double decoupling method for free energy calculation with AMBER, the obtained values for the free energy of water in hydration sites changed from slightly positive up to -5 kcal/mol [3]. The favorable energetic impact of water molecules on complex formation was also found in a study of various protein complexes by Monte Carlo calculations using different force fields [5].

The most important conclusion that can be driven from this free energy analysis is that water molecules in wet spot sites can not be characterized uniformly in energetic terms since in some cases they manifest properties similar to cavity waters and in other do not even contribute favorably to the complex free energy (just occupying an empty space between the residues). Nevertheless, it is realistic to claim that the introduction of water into protein interface description would crucially change the energy function of the system.

4 Conclusions

We present a detailed molecular dynamics study of solvent on 17 protein complexes. Our aim has been to gain insights into the properties of interfacial solvent. We show that water molecules forming wet spots have significantly longer residence time than those on the protein surface, meaning that in terms of mobility interfacial protein residues and interfacial solvent are alike. Although interfacial water molecules are very diverse energetically, their contribution to the free energy of complex formation should be not be ignored. Our data confirm that water plays an important active role in protein interfaces, suggesting that consideration of solvent in the development of energetic functions describing protein interactions is essential.

References

1. T. M. Raschke, *Water structure and interactions with protein surfaces*, Curr. Opin. Struct. Biol. **16**, 152–159, 2006.
2. J. Teyra, A. Doms, M. Schroeder, M. T. Pisabarro, *SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces*, BMC Bioinformatics. **7**, 1004, 2006.
3. D. Hamelberg, J. A. McCammon, *Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method*, J. Am. Chem. Soc. **126**, 7683–7689, 2004.
4. V. A. Makarov, B. K. Andrews, P. E. Smith, B. M. Pettitt, *Residence times of water molecules in the hydration sites of myoglobin*, Biophys. J. **79**, 2966–2974, 2000.
5. C. Barillari, J. Taylor, R. Viner, J. W. Essex, *Classification of water molecules in protein binding sites*, J. Am. Chem. Soc. **129**, 2577-2587, 2007.

Nearly-Deterministic Methods for Optimising Protein Geometry

Gundolf Schenk and Andrew Torda

Zentrum für Bioinformatik, Universität Hamburg, Bundesstraße 43, 20146 Hamburg, Germany
E-mail: {*schenk, torda*}*@zbh.uni-hamburg.de*

Protein structure prediction could be seen as either a challenge or an algorithmic playground. We are certainly interested in algorithmic improvements. Self consistent mean field methods (SCMF) have traditionally been used in areas such as wave function optimisation or protein side-chain placement. We have been trying to apply the ideas to find the most likely conformation for a protein. The philosophy relies on precalculated distributions of structural descriptors given a set of known properties (a protein's sequence). Starting with a sequence, which is decomposed into small overlapping fragments, the conformational space is described by a fixed number of weighted multivariate Gaussians (the known distributions). As the conformational bias, introduced by the sequence fragments, is local the weights of the Gaussians for all overlapping fragments can be optimised iteratively. Unlike molecular dynamics or Monte-Carlo simulations, the optimisation is done in probability space rather than on some initial structure. Therefore, we do not need to calculate energies as in standard SCMF. When the iteration converges sample structures are generated from the weighted Gaussians. The current results show that the procedure is able to find protein-like structures. We can also use this principle to predict protein sequences from structure.

1 Introduction

We are interested in self consistent mean field methods and the protein structure prediction problem. This also means formulating and building new force fields and treating also protein sequence optimisation. Our method has a probabilistic model of protein sequence-structure correlation and approaches self consistency within this framework.

Many methods have already been applied to ab initio protein structure prediction. All use some scoring schemes that are based on statistics and/or physics and chemistry. We want to avoid chemical detail as calculations become intractable and also coarse grain where one is usually dependent on preconceptions. Our approach is purely statistical with its own approximations, but little reliance on human preconceptions.

2 Methods

We have developed and successfully applied a scoring scheme to protein comparisons using sequence, structure or both[1,2]. It is based purely on Bayesian statistics and derived via a maximally parsimonious automatic classifier[3] from overlapping protein fragments. Each is described by 5-7 amino acid types and 10-14 dihedral angles from the backbone. The method assigns a fixed number of class weights (typically 150-300) to each fragment.

With this scoring framework we are able to generate protein structure samples in four steps (figure 1). First, the class weights matrix is build from the sequence. Then, the conformational space is narrowed down by iteratively updating the class weights of overlapping fragments. The local preferences are propagated, as the positions within a fragment

are correlated. This favours the consistent classes. From the conditional class weights sample structures can be generated. As a final step, steric clashes are removed and the models are collapsed by resampling random stretches. Unlike in standard SCMF, the method works without assuming the Boltzmann distribution at any stage.

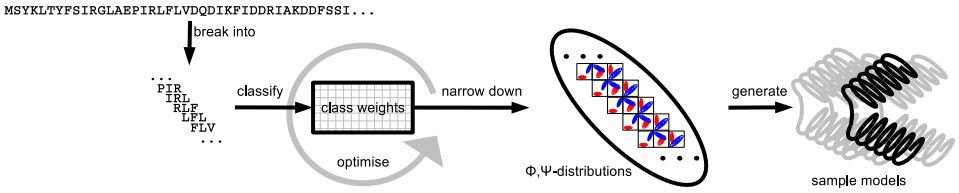


Figure 1. In the online phase a sequence is broken into overlapping fragments. Each fragment is classified leading to a probabilistic description of possible conformations. This can be used to generate sample structures.

The method is available as a web service:
`http://cardigan.zbh.uni-hamburg.de/~mahsch/schenk`

Given an amino acid sequence the server generates a huge number of samples and ranks them with their fragment probabilities,

$$\left(\prod_{i=1}^{N_{\text{fragments}}} \sum_{j=1}^{N_{\text{classes}}} w_{j|f_i^{\text{AA}}} \mathrm{p}_j \left(f_i^{\Phi,\Psi} \right) \right)^{\frac{1}{N_{\text{fragments}}}},$$

where N_{xx} is number of xx, $w_{j|f_i^{\text{AA}}}$ is the conditional weight of class j given the sequence fragment f_i^{AA} and $\mathrm{p}_j \left(f_i^{\Phi,\Psi} \right)$ denotes the conditional density of the structure fragment $f_i^{\Phi,\Psi}$ in class j.

3 Results

The Evaluation of 100000 samples of selected targets (figure 2) suggests that the target structure can be found among the generated models. α-helical targets seem easier than those containing β-strands, which is consistent with other methods. We also calculated the structure which corresponds to the distribution mean. However, we find it far from correct.

There are certain limits with the evaluation one should keep in mind when interpreting the results. The multivariate Gaussian model lacks to account for the periodic nature of dihedral angles and the use of idealised bonds lengths and angles during structure construction introduces a few Ångstrøms error.

Another application of the classification is the prediction of amino acid composition from structure. The regenerated sequences are about 20% identical to the original. So far, it is unknown whether these are bad sequences or alternative possibilities folding to similar structures.

Target Protein				Samples	Best	Highscore	Mean	Mean (opt.)	Samples (opt.)		FB5-HMM*		ROSETTA*	
PDB code	Length	α	β	<6Å [%]	RMSD [Å]	RMSD [Å]	RMSD [Å]	RMSD [Å]	<6Å [%]	RMSD [Å]	<6Å [%]	RMSD [Å]	<6Å [%]	RMSD [Å]
1FC2	43	2	0	3.510	4.1	6.2	6.5	5.5			9.593	2.7	95	3.3
1ENH	54	2	0	0.553	4.4	11.2	9.8	10.7	0.387	4.6	6.595	2.5	47	2.7
2GB1	56	1	4	0.002	5.5	11.3	9.4	9.9	0.001	5.8	0.037	4.9	0	6.3
2CRO	65	5	0	0.050	5.3	8.7	10.6	9.1	0.052	5.2	0.464	3.9	18	4.2
1CTF	68	3	3	0.003	5.6	11.0	12.7	11.0			0.009	5.4	6	5.3
4ICB	76	4	0	0.003	5.7	11.2	10.4	8.0	0.004	5.3	0.089	4.3	17	4.7

Figure 2. Evaluation of 100000 samples of selected targets. *Numbers taken from Ref. 4.

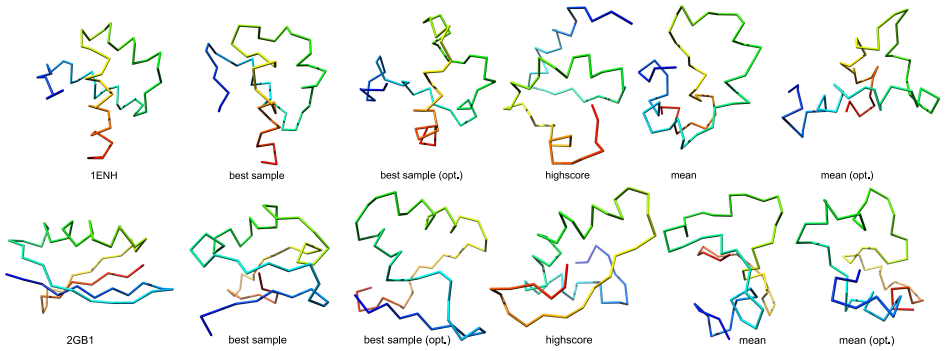


Figure 3. Two Examples from the evaluation.

4 Conclusions and Outlook

The method was tested with only a simple term to favour compaction. It produces the correct type of fold with secondary structure and loops in correct places. There is some limitation of coordinate reconstruction.

We are participating in the CASP8 competition. For this we are improving the server ranking. We are testing a combination with Monte Carlo optimisation methods. To improve the quality of the generated models we are incorporating solvation and long-range terms into our scoring functions. Finally, we are fastening our sampling method.

References

1. G. Schenk, T. Margraf and A. Torda, *Protein sequence and structure alignments within one framework*, Algorithms for Molecular Biology, 3:4, 2008.
2. T. Margraf and A. Torda, *Salami Server*, `http://www.zbh.uni-hamburg.de/salami`.
3. P. Cheeseman and J. Stutz, *Bayesian Classification (AutoClass): Theory and Results*, Advances in Knowledge Discovery and Data Mining, 1996.
4. T. Hamelryck, J. T. Kent and A. Krogh, *Sampling Realistic Protein Conformations Using Local Structural Bias*, PloS Comput Biol **2(9):e131**, 1121–1133, 2006.

Different Types of Protein Folding Identified with a Coarse-Grained Heteropolymer Model

Stefan Schnabel, Michael Bachmann, and Wolfhard Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ), Universität Leipzig,
Postfach 100 920, D-04009 Leipzig, Germany
E-mail: {schnabel, bachmann, janke}@itp.uni-leipzig.de

Applying multicanonical simulations we investigated folding properties of off-lattice heteropolymers employing a mesoscopic hydrophobic-polar model. We study for various sequences folding channels in the free-energy landscape by comparing the equilibrium conformations with the folded state in terms of an angular overlap parameter. Although all investigated heteropolymer sequences contain the same content of hydrophobic and polar monomers, our analysis of the folding channels reveals a variety of characteristic folding behaviors known from realistic peptides.

1 Introduction

The identification of folding channels is one of the key tasks of protein folding studies. While secondary structures depend on atomistic details (such as, in particular, hydrogen bonding), tertiary structure formation should exhibit a certain degree of universality. This suggests that coarse-grained models might capture the main characteristics on mesoscopic scales of this stage of the folding process. In this note we report a computer simulation study that tests this idea by employing the off-lattice hydrophobic-polar AB model.

2 Model and Method

In the AB model[1] a heteropolymer or coarse-grained peptide is described as a chain of hydrophobic (A) and hydrophilic (B) monomers whose energy is obtained from specific Lennard-Jones potentials between all nonbonded pairs. Thereby an aqueous environment is modeled implicitly by the energetically favored A-A contacts that lead to the formation of a hydrophobic core at low temperatures. A smaller energy contribution arises from the local bending of the chain.

In the Monte Carlo simulations we applied the multicanonical technique[2] where an additional weight function leads to a flat distribution in energy space. This allows the reweighting of the data to any desired temperature with equally high accuracy. In order to identify folding channels we used the angular overlap parameter Q introduced in Ref. 3 to compare equilibrium conformations at temperature T with the previously determined ground-state structure. Two conformations are equal if $Q = 1$. The temperature-dependent probability distributions of the total energy E and Q, $P_T(E, Q)$, then allow the analysis of the chain's folding behavior (for more details, see Ref. 4).

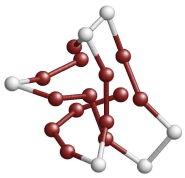


Figure 1. Ground-state conformation of $A_4BA_2BABA_2B_2A_3BA_2$.

3 Results

Folding through intermediates: For the sequence $A_4BA_2BABA_2B_2A_3BA_2$ the conformations at higher temperatures do not exhibit significant similarities ($Q \approx 0.7$) with the ground-state conformation depicted in Fig. 1. As the distributions $P_T(E,Q)$ in Fig. 2 show, the main branch slightly moves to higher Q-values with decreasing temperature until it splits below $T \approx 0.2$. Near $T \approx 0.1$, the population of intermediate conformations has increased and coexists with denatured states. Approaching $T \approx 0.05$, the intermediate states dominate the canonical ensemble. The probability for denatured conformations is reduced, but the onset of an occupation of states with similarities to the global-energy minimum (GEM) is clearly visible. At $T = 0.02$, the majority of conformations has a large overlap with the ground state and the heteropolymer has folded into its native state.

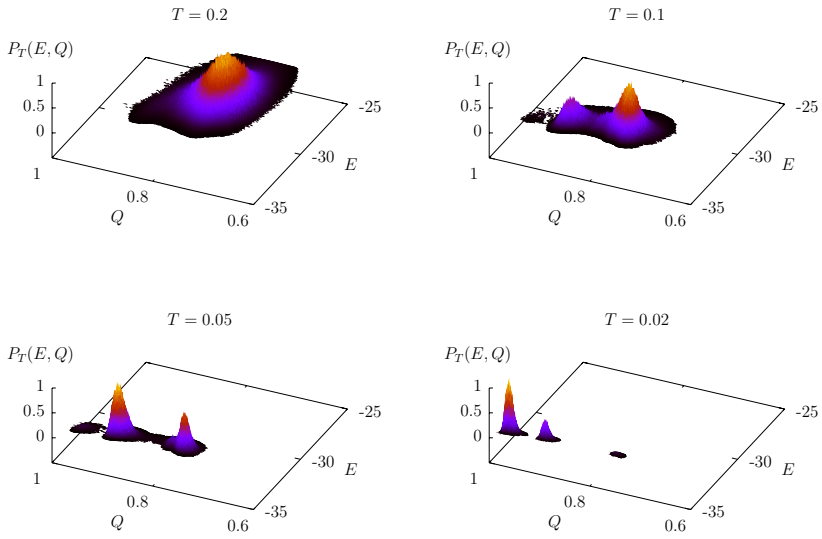


Figure 2. Canonical distributions $P_T(E,Q)$ of energy E and overlap Q with the GEM for the sequence $A_4BA_2BABA_2B_2A_3BA_2$.

Two-state folding and metastability: A typical two-state folding scenario can be observed for the sequence $BA_6BA_4BA_2BA_2B_2$, where again $Q \approx 0.7$ at high temperatures indicates that the heteropolymer is in a random state that possesses no similarities with the GEM conformation. Around $T = 0.1$ the canonical ensemble divides and a state with $Q > 0.9$ is occupied in addition to the disordered conformations with $Q < 0.8$. With decreasing temperature the newly formed branch dominates the ensemble and approaches $Q = 1.0$ when $T \to 0$.

An example for metastability is provided by a third heteropolymer with the sequence $A_4B_2A_4BA_2BA_3B_2$, which exhibits in the low-temperature phase a glassy behavior. While at high temperature similar to the other cases all conformations are distributed around $Q \approx 0.7$, the main folding channel does not lead to a single ground-state conformation. Instead two rival conformations ($Q \approx 1$ and $Q \approx 0.75$) can be found also at temperatures below $T = 0.01$.

4 Summary

Numerical investigations of medium size or large proteins by means of Monte Carlo simulations are very difficult since complex energy functions demand high computational efforts. On the other hand, problems of comparable complexity with equal or higher number of degrees of freedom can be handled using simplified coarse-grained models like the AB model. Although the model is relatively simple and the investigated sequences are only permutations of each other (and were not designed especially), three different kinds of folding could be observed. Since all of them – folding through intermediates, two-state folding and metastability – are also known from real peptides, the AB model seems to resemble general characteristics of protein folding.[4] Therefore we believe that further research on this model offers the possibility to gain qualitative insights in tertiary folding where microscopic details are expected to be of less importance.

Acknowledgments

This work is supported by the DFG (German Science Foundation) under Grant No. JA 483/24-1/2, the Graduate School "BuildMoNa" and the DFH-UFA PhD College CDFA-02-07. Some simulations were performed on the supercomputer JUMP of the John von Neumann Institute for Computing (NIC), Forschungszentrum Jülich under Grant No. hlz11.

References

1. F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, Phys. Rev. E, **48**, 1469, 1993.
2. B. A. Berg and T. Neuhaus, Phys. Lett. B, **267**, 249, 1991.
3. M. Bachmann, H. Arkın, and W. Janke, Phys. Rev. E, **71**, 031906, 2005.
4. S. Schnabel, M. Bachmann, and W. Janke, Phys. Rev. Lett. **98**, 048103, 2007; J. Chem. Phys. **126**, 105102, 2007.

Effect of Surfaces on the Aggregation of Hydrophobic and Hydrophilic Amyloidogenic Peptides

Gurpreet Singh, Ivan Brovchenko, and Roland Winter

Department of Chemistry, Biophysical Chemistry, Dortmund University of Technology, D-44227 Dortmund, Germany
E-mail: {*gurpreet.singh, ivan.brovchenko, roland.winter*}*@tu-dortmund.de*

The general effect of surface hydrophobicity/hydrophilicity on the aggregation of peptides is studied by simulations of oversaturated aqueous solutions of hydrophobic and hydrophilic peptides in pores with hydrophobic (paraffin-like) and hydrophilic (silica-like) walls. Strong adsorption of peptides on the pore walls is observed in the case of the hydrophobic peptides in a hydrophobic pore, where all peptides are strongly adsorbed and aligned parallel to the walls already after 30 ns. Adsorption of this peptide at the liquid-vapor interface is quite similar. In the other three cases considered, the peptides are repelled from the walls, localized near the pore center and do not show orientational ordering with respect to the walls. Our results show that even a single factor such as the water density distribution has a drastic effect on the character of peptide aggregation near surfaces. A wider diversity of possible scenarios can be expected when specific peptide-surface interactions are taken into account.

Adsorption of peptides on surfaces can strongly affect their aggregation. Possible enhancement of the orientational ordering of peptides near surfaces can promote the formation of ordered peptide aggregates. This may be one of the factor, which makes the intracellular and extracellular aggregation different. To explore the general effect of surface hydrophobicity/hydrophilicity on peptide adsorption and aggregation, we performed a series of computer simulation studies on oversaturated aqueous solutions of peptides in slit-like pores with smooth walls interacting via a (9-3) LJ potential with water molecules and non-interacting with peptides. Two kinds of amyloidogenic peptides were used: the hydrophobic peptide NFGAIL, (residues 22-27 of the human islet amyloid polypeptide), and the polar hydrophilic peptide GNNQQNY (residues 7-13 of the yeast prion Sup35).

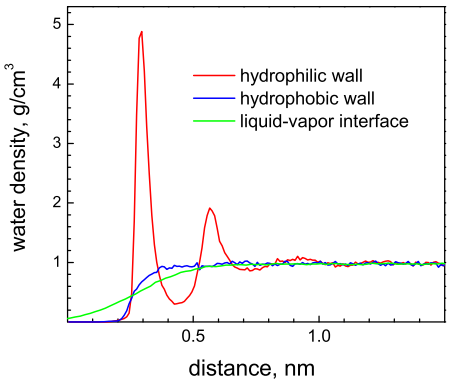


Figure 1. Water density profiles near hydrophobic and hydrophilic walls and at the liquid-vapor interface.

Two kinds of pore walls were considered: a hydrophobic paraffin-like wall, which causes a pronounced water density depletion near the surface (see the blue line in Fig. 1), and a hydrophilic silica-like wall, which causes formation of two highly ordered water layers near the surface (see the red line in Fig. 1). Additionally, we have simulated a liquid-vapor interface of the aqueous solution of hydrophobic peptides (its density profile is shown by the green line in Fig. 1).

Six peptide fragments were randomly inserted in a cubic box of length 6 nm such that the peptides are at least 0.7 nm away from each other and 0.15 nm away from the surfaces. All atomic molecular dynamic simulations were carried out at T = 330 K with Gromacs software using the OPLS force field and SPCE water molecules.[1] The PME method was used to treat long range electrostatic interactions.[2] Five simulations with different initial velocities were carried out for the duration of 70 ns for each type of surface and peptide combination.

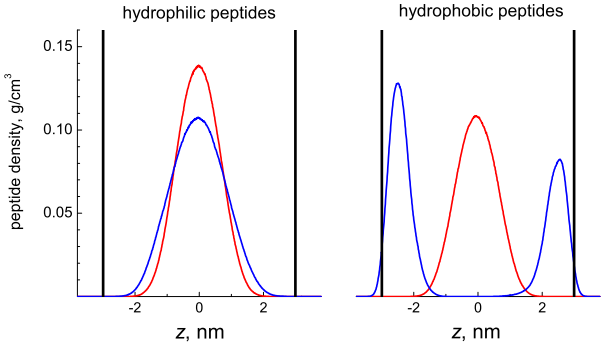


Figure 2. Density profiles of hydrophilic (left) and hydrophobic (right) peptides in the pores with hydrophilic (red) and hydrophobic (blue) walls. Vertical black lines indicate locations of the pore walls. The pore center is located at z = 0.

The density profiles of the peptide in pores, calculated by taking into account all peptide atoms, are shown in Fig. 2. Hydrophilic peptides show strong desorption from both hydrophilic and hydrophobic surfaces and concentrate near the pore center (see left panel in Fig. 2). The repulsion of these peptides is slightly stronger in the case of a hydrophilic surface. Hydrophobic peptides exibits a quite similar desorption characteristics from a hydrophilic surface (see red line in the right panel in Fig. 2). An opposite behavior is observed for the hydrophobic peptides in hydrophobic pore: in all simulation runs, the peptides are strongly adsorbed at the walls already after 20 to 30 ns (see blue line in the right panel in Fig. 2). As the peptides do not interact with the pore walls, the tail of peptide's density profile crosses the pore wall. Quite similarly, hydrophobic peptides strongly adsorb at the liquid-vapor interface (not shown), which also has a strong hydrophobic character.

A strong adsorption of the peptides at the surface enhances their orientational ordering. The probability distribution of the angle α between the pore surface and the vector connecting two most distant peptide heavy atoms is shown in Fig. 3. When peptides are repelled from the pore walls and localized in its center, the orientation of their longest axis is highly isotropic (left panel in Fig. 3). In contrast, strong adsorption of the peptides at

the surface causes their longest axis aligned parallel to the walls. Thus, adsorption at the surface not only speeds up their aggregation, but also provides favourable conditions for the formation of ordered peptide aggregate characterized by extensive β-sheet formation.

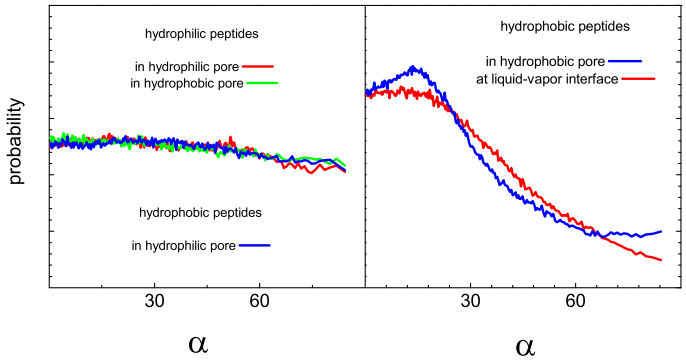


Figure 3. Probability distribution of the angle α between the pore surface and the vector connecting two most distant peptide heavy atoms.

The results shown in Figs. 2 and 3 evidence that even a single factor such as the water density distribution near a surface has a drastic influence on the peptide aggregation. Hence, the surface effects are expected to be much more important in the case of intracellular peptide aggregation. Presumably, such surface effects and the effect of the finite size of biological cells[3] are the two main factors rendering the intracellular and extracellular peptide aggregation processes different.

Acknowledgments

Financial support from the International Max-Planck Research School in Chemical Biology and from the country NRW and the EU (Europäischer Fonds für regionale Entwicklung) is gratefully acknowledged.

References

1. David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen, *GROMACS: fast, flexible, and free.*, J Comput Chem, **26**, no. 16, 1701–1718, Dec 2005.
2. ln Chul Yeh and Max L. Berkowitz, *Ewald summation for systems with slab geometry*, J. Chem. Phys., **111**, 3155, 1999.
3. Gurpreet Singh, Ivan Brovchenko, Alla Oleinikova, and Roland Winter, *Aggregation of Fragments of the Islet Amyloid Polypeptide as a Phase Transition: A Cluster Analysis*, in: From Computational Biophysics to Systems Biology (CBSB07), Proceedings of the NIC Workshop 2007, Ulrich H. E. Hansmann, (Ed.), vol. 36, pp. 275– 278, 2007.

Free Energy Study of Ion Permeation through Gramicidin

Shirley W. I. Siu and Rainer A. Böckmann

Theoretical & Computational Membrane Biology, Center for Bioinformatics, Saarland University,
P.O. Box 15 11 50, 66041 Saarbrücken, Germany
E-mail: {siu,rainer}@bioinformatik.uni-saarland.de

The pentadecapeptide gramicidin forms a cation-specific ion channel in membrane environment. Two conformations are known up-to-date: the head-to-head helical dimer (HD) and the intertwined double helical form (DH). These two conformations are favored depending on the specific conditions, but the biologically active form is still a matter of debate. Nevertheless, due to its small size, the gramicidin serves as an excellent ion channel model for both computational and experimental studies. In this comparative study, we focus on the energetics of single potassium ion permeation by means of the potential of mean force (PMF) for both gramicidin conformations using molecular dynamics simulations. Our results show that the DH has a significantly decreased central barrier with respect to HD, implying an increased ion conduction. The barrier to ion passage is found to be closely related to the channel flexibility. Multiple ion permeation for the DH conformation is probably facilitated due to its opposing pore water dipole moments at the pore entrances.

1 Introduction

Gramicidin A (gA) is the major component of the antibiotic gramicidin from the soil bacteria *Bacillus brevis*. Each monomer is made up of 15 alternating L- and D-amino acids capped at the two ends by a formyl group and an ethanolamine group. When dimerized, gA functions as a cation-selective transmembrane channel. The unique sequence of the gA peptide is able to adopt a wide range of conformations based on various environmental factors. Mainly, two folding motifs of gramicidin were reported in structural experimental studies, namely the single-stranded head-to-head dimer (HD) and the double-stranded helical dimer (DH). While the head-to-head dimer was believed to be the more thermodynamically stable form in the membrane, experiments demonstrated that the double-stranded dimer coexists with the single-stranded form in the membrane in certain proportion[1].

Due to its small size and well-defined channel pore, gramicidin is a popular model for studying the properties and mechanism of ion conduction. Several computational studies, mainly of the HD dimer were published recently[2]. All of them showed that the free energy of ion permeation through the HD channel contains an unexpectedly high central energy barrier and relatively weak binding sites at the two mouths of the channel. On the other hand, the DH conformation has been shown to translocate a water column at an increased rate as compared to HD[3,4]. The opposing pore water dipoles found only in DH suggested a facilitation of multiple ion passage in the double-stranded structure.[4]

In this study, we focus on the ion conduction properties of both DH and HD conformations. By employing the free energy calculation method, we construct the potential of mean force (PMF) of the ion permeation pathway. Our result show a decreased free energy barrier and an increased structural flexibility for DH as compared to HD.

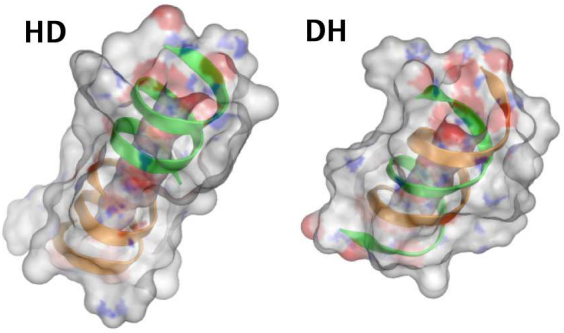


Figure 1. The two major conformations of gramicidin (PDB 1MAG for HD, 1AV2 for DH) drawn in its solvent accessible surface (water radius 1.4 Å).

2 Method

The simulation system consists of a gA (see Figure 1), 124 dimyristoylphosphatidylcholine (DMPC) lipids and 6,142 water molecules with an ionic concentration of 200 mM KCL. For the PMF of ion permeation, the ion's positions along the gramicidin channel are sampled applying the umbrella sampling technique. The PMF is then calculated by unbiasing and combining the ion density distributions of the window simulations (0.5 Å intervals) along the direction of the bilayer normal using the Weighted Histogram Analysis Method (WHAM). Molecular dynamics simulations were performed using the GROMACS package with the GROMOS53a6 protein force field and the Berger lipid force field.

3 Results

The symmetrized PMF profiles of the two gramicidin conformations show a remarkable difference in stabilizing a K^+ ion along the gramicidin channel. As shown in Figure 2, the HD profile has a large central barrier of about 46 kJ/mol. A wide shallow well is observed at the interface of the channel and the lipid headgroup region. Ion entering the channel experience a stepwise increase in free energy. Results obtained in this study for the HD conformation are in good agreement with previous free energy calculations for the barrier height[2]. In contrast, the DH profile has a much less rugged energy profile with a decreased central barrier of only 15 kJ/mol, at least a factor of three lower than for HD. Binding sites at the channel entrance as well as in the lipid interfacial region are clearly seen in the symmetrized profile. Experimentally, it was observed that gramicidin contains two symmetrically related binding sites at both ends of the channel.

Structural changes of the channel in response to the presence of a K^+ ion are e.g. reflected by changes in the pore radius. The DH conformation has a pore with a uniform radius of ≈ 1.7 Å in the ion-free state. In contrast, the pore is more narrow for the HD conformation at the channel entrances (≈ 1.0 Å). In the window simulations, drastic changes close to the channel openings were observed in DH which amounts to a reduction of about 30% in the pore radius in contrast to only 20% in HD. This reflects the high flexibility of the double-helical conformation in response to the conducting ion.

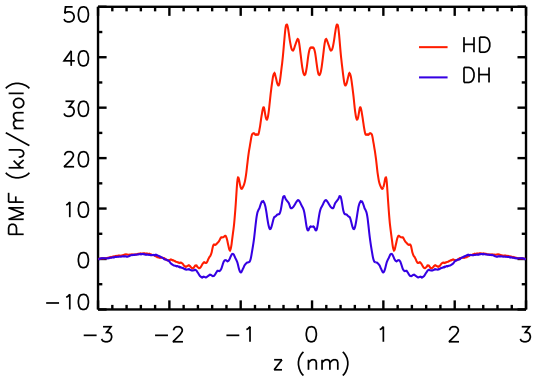


Figure 2. Compare the PMF profiles of K^+ permeating the gramicidin channel in HD and DH conformation (symmetrized).

4 Conclusion

We have shown that the double-helical dimer of gramicidin has a 3-fold decreased free energy barrier for ion permeation compared to the single-stranded dimer, possibly coupled to the high flexibility of the DH channel in coordinating the passing ion.

Acknowledgments

The WHAM program was kindly provided by B. de Groot from MPI for Biophysical Chemistry in Göttingen. Financial support by the Deutsche Forschungsgemeinschaft (Graduate School *Structure Formation and Transport in Complex Systems* No. 1276/1) is acknowledged. As members of the Center for Bioinformatics, the authors are supported by the Deutsche Forschungsgemeinschaft BIZ 4/1.

References

1. M. C. Bano, L. Braco, and C. Abad, *Conformational transitions of gramicidin A in phospholipid model membranes. A high-performance liquid chromatography assessment*, Biochemistry **30**, 886–894, 1991.
2. T. W. Allen, O. S. Andersen, and B. Roux, *Ion permeation through a narrow channel: Using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields*, Biophys. J. **90**, 3447–3468, 2006.
3. B. L. de Groot, D. P. Tieleman, P. Pohl, and H. Grubmüller, *Water Permeation through Gramicidin A: Desformylation and the Double helix: A Molecular Dynamics Study*, Biophys. J. **82**, 2934–2942, 2002.
4. S. W. I. Siu and R. A. Böckmann, *Electric field effects on membranes: Gramicidin A as a test ground*, J. Struct. Biol. **157**, 545–556, 2007.

Free-Energy Based All-Atom Protein Folding Using Worldwide Distributed Computational Resources

Timo Strunk[1,2]**, Srinivasa Murthy Gopal**[1]**,**
Konstantin Klenin[1]**, and Wolfgang Wenzel**[1]

[1] Institute of Nanotechnology, Research Centre Karlsruhe,
76344 Eggenstein-Leopoldshafen, Germany
E-mail: {Timo.Strunk, wenzel}@int.fzk.de

[2] Fachbereich Physik, Technische Universität Dortmund,
Otto-Hahn-Str. 4, 44227 Dortmund, Germany

We have implemented massively parallel stochastic optimization methods for all-atom de-novo protein folding using our free-energy forcefield PFF02[1,2], which is based on Anfinsen's thermodynamic hypothesis[3]. We have implemented this approach (POEM) using a world-wide volunteer computational grid to predictively and reproducibly fold the HIV accessory protein 1F4I from completely unfolded conformations.

1 Motivation

Protein folding and structure prediction at the all-atom level remain important computational challenges. To achieve this goal in the long term it is important to develop methods that are capable to predictively fold proteins and peptides from unbiased unstructured conformations to the native ensemble. Direct simulation studies have demonstrated the folding of several small peptides and mini-proteins from completely extended conformations, but remain limited in the system size by the large computational effort required.
One great hope towards reproducible all-atom folding is the development of algorithms that can exploit emerging massively parallel computational architectures.We have recently developed an evolutionary algorithm, which generalized the basin hopping or Monte-Carlo with minimization[4,5], method to many concurrent simulations. This approach was ported to the massively parallel BOINC architecture on POEM@HOME (`http://boinc.fzk.de`) and verified by folding several proteins reproducibly.

2 Method

We have parameterized an all-atom free-energy forcefield for proteins (PFF01/02)[1,2], which is based on the fundamental biophysical interactions that govern the folding process. We could show that near-native conformations of several proteins correspond to the global optimum of this forcefield. We have also developed, or specifically adapted, efficient stochastic optimization methods (stochastic tunnelling, basin hopping, evolutionary algorithms) to simulate the protein folding process. Forcefield and simulation methods are implemented in the POEM (Protein Optimization with free-Energy Methods) program package.

2.1 Optimization Strategy

We have generalized this method to a population of fixed size which is iteratively improved by an arbitrary number of concurrent dynamical processes[6,7]. The whole population is guided towards the optimum of the free energy surface with a simple evolutionary strategy in which members of the population are drawn and then subjected to a single simulated annealing basin hopping cycle. At the end of each cycle the resulting conformation either replaces a member of the active population or is discarded. The decision tree for this process is illustrated in figure 1.

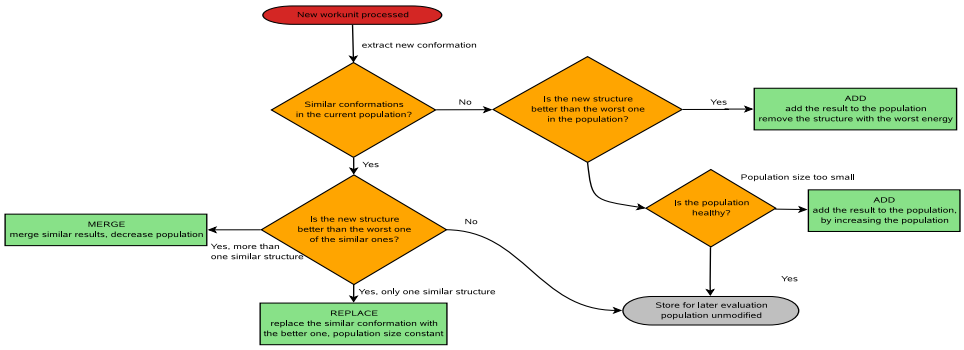


Figure 1. Evolutionary optimization strategy.

2.2 POEM@HOME

In this investigation, we deployed the BOINC server POEM@HOME (`http://boinc.fzk.de`), which explores the free-energy landscape in many parallel dynamical processes, which are coordinated in a single evolutionary algorithm population as outlined in section 2.1. The overall computational work is thus segmented into medium size work-units, which can be processed independently.

3 Results

3.1 Folding of the HIV Accessory Protein 1F4I

The 40 amino acid target 1F4I was folded using the evolutionary algorithm on POEM@HOME. The population was initially seeded with a single extended 'stick' conformation. Figure 2 shows the convergence of the energy as a function of the total number of basin hopping cycles. We find that the best energy converges quickly to a near-optimal value with the total number of basin hopping cycles. As a result of the population diversity criterion, there will always be a finite difference between the average and best energy. This is established by an acceptance threshold of 3 Å RMSD for the inclusion of new structures into the population.

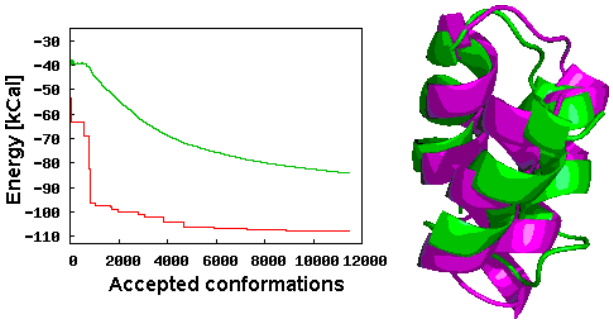


Figure 2. Left: Evolution of average (upper curve) and best (lower curve) energy in the folding process. Right: Overlay of the simulated and experimental structure of 1F4I.

E [$kcal$]	RMSD [Å]	Secondary Structure
Exp	-	CCHHHHHHHHHTTCCHHHHHHHHHTTTSCSHHHHHHHHHC
-107.77	2.56	CHHHHHHHHHHSCCHHHHHHHHHHHHHCHHHHHHHHHC
-107.12	8.11	CHHHHHHHHHSCCSSSSCBTTSCCSHHHHHHHHSCSBC
-106.30	6.60	CHHHHHHHHHHCSSSHHHHHHHHHHHHCHHHHHHHHHC
-103.90	7.95	CCHHHHHHHHSCCSSSEEBTTBCSSHHHHHHHHCCEEC
-103.66	4.90	CCHHHHHHHHHCCCHHHHHSCCBTTTBHHHHHHHHHC

Table 1. Top five best energy structure of different topology for folding Target 1F4I.

Another indication of the diversity of the algorithm can be observed in table 1. All of the best energy structures show significant differences in their secondary structures. The best energy-structure found has a RMSD of 2.56Å to the native structure.

4 Conclusions

We have shown, that the mapping of the 'folding problem' onto an optimization problem permits the use of methods that speed the exploration of the free-energy surface. The present study demonstrates, equally importantly, that it is possible to parallelize the search process by splitting the simulation into a large number of independent conformations, rather than by parallelizing the energy evaluation.

Acknowledgments

We thank the Fond der Chemischen Industrie, the BMBF, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-1, WE 1863/10-2 and WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. We also thank the BOINC (`http://boinc.berkeley.edu`) developers and especially our current POEM@HOME Users for their support.

References

1. A. Verma, W. Wenzel, *All-atom protein folding using an improved free-energy force-field*, Biophys. J: Suppl. **S**, 214, 2007.
2. T.Herges, W. Wenzel, *An all-atom forcefield for teriary structure prediction of helical proteins*, Biophys. J **87**, 3100, 2004.
3. C. B. Anfinsen, *Principles that govern the folding of protein chains*, Science **181**, 223–230, 1973.
4. T.Herges, W.Wenzel, *In Silico Folding of a Three Helix Protein and Characterization of Its Free-Energy Landscape in an All-Atom Force Field*, Phys.Rev.Lett. **94**, 018101, 2005.
5. A.Verma, A.Schug, K.H. Lee, W.Wenzel, *Basin hopping simulations for all-atom protein folding*, J.Chem.Phys. **124**, 044515, 2006.
6. A.Schug, W.Wenzel., *Predictive in-silico all-atom folding of a four helix protein with a free-energy model*, J. Am. Chem. Soc. **126**, 16736–16737, 2004.
7. A.Schug, W.Wenzel., *An evolutionary strategy for all-atom folding of the sixty amino acid bacterial ribosomal proein L20*, Biophys. J. **90**, 4273–4280, 2006.

Parallel 2d-Wavelet Transform on the Cell/B.E. for Fast Calculation of Coulomb Potentials

Annika Schiller and Godehard Sutmann

Institute for Advanced Simulation, Jülich Supercomputing Centre,
Research Centre Jülich, 52425 Jülich, Germany
E-mail: {a.schiller, g.sutmann}@fz-juelich.de

1 Introduction

Long range interactions between particles often play an important role in biomolecular simulations in order to describe the structure and dynamics of particles correctly. The calculation of this type of interaction often limits the time and length scale of a simulation, since it scales as $\mathcal{O}(N^2)$, where N is the number of particles in the system. In order to overcome this limitation, different types of fast algorithms of order $\mathcal{O}(N \log N)$ or $\mathcal{O}(N)$ were developed (for an overview see e.g. Ref. 1, 2). One of this type of algorithms is based on a Wavelet transform technique[3]. A computationally intensive part consists in preparing the 2d-Wavelet transform of inverse distances between fixed grid points in space, onto which particle properties are transferred. Since the grid is static, this computational intense part has only to be performed once. Due to memory requirements and performance, it is desirable to perform these calculations on a scalable computer architecture. To this end the Cell Broadband Engine (Cell/B.E.) heterogeneous multicore processor was chosen to explore its capabilities and high potential in performance. The processor characteristics include multiple heterogeneous execution units, SIMD processing engines, fast local store and a software managed cache. Applications can achieve a performance which is close to the theoretical peak performance if specific features are respected.

In the present work an implementation of the fast 2d-Wavelet transform, realized via a triple matrix multiplication, was developed using the native programming API of the IBM Software Development Kit (CellSDK[4]). In this implementation the architectural requirements of the Cell processor are taken into account and Cell specific optimizations are applied where practical. Therewith, the difficulties and problems in porting code to Cell/B.E. and using sparse linear algebra operations on the Cell processor are assessed. Finally the results of the native implementation are discussed and compared to the results of the same algorithm implemented via the Cell Superscalar framework (CellSs[5]), a high-level portable programming model.

2 Method

The method of calculating Coulomb potentials with the help of Wavelets was described in Ref. 3. Here, we concentrate on the description on how to calculate the triple-matrix-multiply

$$\tilde{\mathbf{A}}_t = \mathbb{T}\{\mathbf{W}_l \, \mathbf{A} \, \mathbf{W}_l^T; \, t\} \tag{1}$$

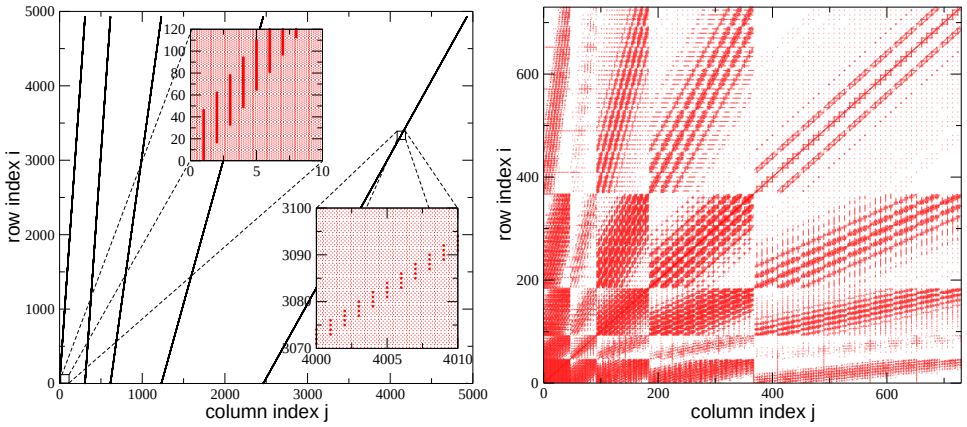


Figure 1. Structure of matrices $\mathbf{W}_l$ (left) for a Daubechies-4 Wavelet-basis (details are shown in the insets) and $\tilde{\mathbf{A}}_t$ (right) with threshold value $t = 0.01$ and compression rate $\chi = 81\%$.

where the matrix $\mathbf{W}_l$ is the resulting Wavelet transform matrix on level l, which is sparse and contains both the Wavelet and scaling coefficients. The structure of this matrix is displayed in Fig.1. On the other hand the matrix $\mathbf{A} \in \mathbb{R}^{N_g \times N_g}$ is dense, as it contains the inverse distances between equi-spaced grid points, where N_g is the number of grid points in the system. Since $N_g \propto N$, the dimension of the matrix may become rather large, which requires an economic memory management. Finally, $\tilde{\mathbf{A}}_t$ is a thresholded Wavelet transform of $\mathbf{A}$, which results from the threshold operation $\mathbb{T}\{.\,;\,t\}$, where all absolute values below a given value t are set to zero.

In order to apply the method also to larger problem sizes, the Wavelet matrix $\mathbf{W}_l$ is partitioned and transferred in parts to the SPU. Information about the matrix is stored in Compressed Sparse Row (CSR) format[6], for which four arrays are introduced: (i) the elements of the matrix, `j_elem`; (ii) column index of n-th element, `j_index`; (iii) offset value, `nnn_off`, in order to address the first element of row i; (iv) number of elements of row i, `nj_w`.

The partitioning of $\mathbf{W}_l$ does not guarantee that every block of data contains complete rows. Transferring incomplete rows to the SPU would, however, increase the complexity of the computation algorithm. A possible method to avoid splitting of rows consists in padding. Therefore, in order to align the array `j_elem` a second array has to be allocated where rows of $\mathbf{W}_l$ are grouped into 16 kB blocks. If one row is split by a block border the complete row is put into the next block and the current block is filled by padding. The handling of array `j_index` proceeds analogously. The rows of $\mathbf{W}_l$ do not always contain the same number of non-zeros. Therefore, an array is introduced which stores the number of rows in each block. Using this block information the corresponding elements of arrays `nnn_off` and `nj_w` can be loaded. Note, that these arrays are not aligned. Therefore it is necessary to expand the needed part to an aligned extract which contains all the data actually required.

In the present implementation, the matrix $\tilde{\mathbf{A}}_t$ is calculated column by column in two

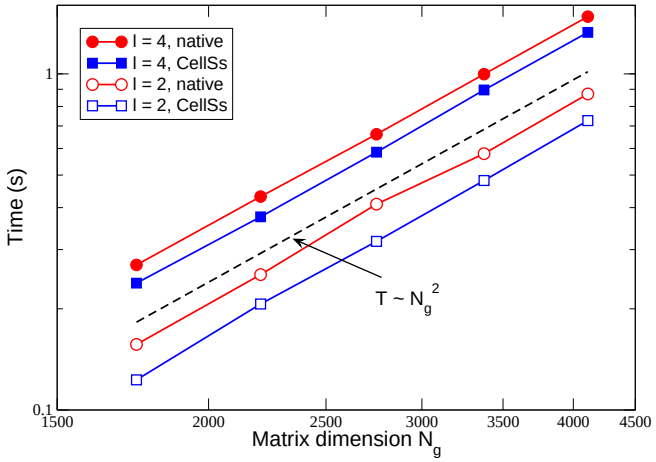


Figure 2. Timings for a native implementation of the algorithm and for the CellSs implementation for Wavelet transforms of different level l. CellSs outperforms the native implementation, since features like double-buffering are not considered in the native version.

steps as follows[7] (for different implementations cf. Ref. 8)

$$\tilde{\mathbf{A}} \;= \mathbf{W}_l \, \mathbf{A} \, \mathbf{W}_l^T \;\; \Rightarrow \;\; \tilde{\boldsymbol{a}}_q = \mathbf{W}_l \times (\mathbf{A} \times \boldsymbol{w}_q^T) \tag{2}$$

In this equation $\tilde{\boldsymbol{a}}_q$ is the q-th column of the result matrix $\tilde{\mathbf{A}}$ and $\boldsymbol{w}_q^T$ is the q-th column of the transposed Wavelet matrix $\mathbf{W}_l^T$. In the first step the intermediate result $\boldsymbol{b}_q = \mathbf{A} \times \boldsymbol{w}_q^T$ is calculated. In the second step the q-th column of $\tilde{\mathbf{A}}$ is calculated via $\tilde{\boldsymbol{a}}_q = \mathbf{W}_l \times \boldsymbol{b}_q$. These steps have to be repeated for each of the N columns. In the implementation of this algorithm the intermediate result $\boldsymbol{b}_q$ is buffered in a temporary array on the SPU. To decrease DMA transfers this intermediate array is not transferred back to the PPU.

The two main aspects which are to be considered for this application are the amount of data and the number of floating point operations on the SPU. Since for realistic applications $N_g \approx 32^3$, the size of $\mathbf{A}$ easily exceeds the memory capacities of the SPU (256 kB). However, considering the inverse distance matrix in detail, reveals that full storage of elements contains a lot of redundancy. If $N_g = n^3$ there are $N_g(N_g - 1)$ distances between grid points, of which only $n(n+1)(n+2)/6 - 1$ are different. Therefore, only non-redundant data are calculated on the PPU and transferred to the SPU, where an addressing of matrix elements is performed in order to map a three-dimensional grid onto a two-dimensional matrix. That means that it is enough to store $\mathcal{O}(N_g)$ values for the kernel of $\mathbf{A}$ instead of $\mathcal{O}(N_g^2)$ because of redundant entries. The Wavelet matrix $\mathbf{W}_l$ is blocked and loaded blockwise onto the SPUs, where $\tilde{\mathbf{A}}$ is then calculated in two steps column by column and then thresholded. The intermediate result is buffered on the SPU and not transferred back to the PPU.

3 Results

A native programming approach was compared to one, where CellSs was used. Fig. 2 shows the result of this comparison for Wavelet transforms of different sizes N_g and levels l for a Haar-Wavelet transform.

As a first result it is found that the scaling of the algorithm is $\mathcal{O}(N_g^2)$, as it could be expected for a sparse-dense-sparse matrix multiply. This result, however, shows that (i) the algorithm is efficiently implemented and (ii) with increasing problem size the PPU-SPU-communication is not a limiting factor. This leads to optimism for increasing problem sizes.

As a second result it is seen that the CellSs based implementation outperforms the native implementation in all cases by about 10-20%. This prooves that CellSs is not only a very nice tool to ease Cell programming, but that it is also able to produce high performance code. The reason for the difference of the two approaches may be found in neglecting SIMD vectorization and features like double buffering in the native implementation. Since the complexity of the code already was large, these features were postponed in first instance. The result shows that CellSs leads to very good performance while reducing the programming effort at the same time.

References

1. P. Gibbon and G. Sutmann. Long range interactions in many-particle simulation. In J. Grotendorst, D. Marx, and A. Muramatsu, editors, *Quantum simulations of many-body systems: from theory to algorithms*, volume 10, pages 467–506, Jülich, 2001. John von Neumann Institute for Computing.
2. M. Griebel, S. Knapek, and G. Zumbusch. *Numerical Simulation in Molecular Dynamics*. Springer, Berlin, 2007.
3. G. Sutmann and S. Wädow. A Fast Wavelet Based Evaluation of Coulomb Potentials in Molecular Systems. In U.H.E. Hansmann, editor, *From Computational Biophysics to Systems Biology 2006*, volume 34, pages 185–189, Jülich, 2006. John von Neumann Institute for Computing.
4. IBM, *Cell Broadband Engine: Programming Tutorial*, March 2007, Version 2.1.
5. `http://www-03.ibm.com/technology/cell/swlib.html`.
6. Y. Saad. SPARSKIT: A Basic Tool for Sparse Matrix Computation. Technical Report CSRD TR 1029, University of Illinois, 1990.
7. F. Marino, V. Piuri, and E.E. Swartzlander. A parallel implementation of the 2d discrete Wavelet transform without interprocessor communications. *IEEE Trans. Sign. Proc.*, 47:3179–3184, 1999.
8. L. Yang, A. Schiller, G. Sutmann, B.J.N. Wylie, R. Altenfeld, and F. Wolf. Comparison of CellSs and native programming with a Jacobi solver and triple-matrix-multiply on Cell/B.E. Proc. of the *9th International Workshop on State-of-the-Art in Scientific and Parallel Computing* Para 2008 (submitted).

RedMD – A New Package for Reduced Molecular Dynamics

Marcin Szypowski[1], **Adam Górecki**[1,2], **and Joanna Trylska**[1]

[1] Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland
E-mail: {M.Szypowski, joanna}@icm.edu.pl

[2] Department of Biophysics, Faculty of Physics, University of Warsaw, Warsaw, Poland
E-mail: A.Gorecki@icm.edu.pl

We developed a RedMD package to perform molecular dynamics simulations for coarse-grained models of proteins, nucleic acids and its complexes. Simulations can be carried out in the microcanonical ensemble, as well as with Berendsen and Langevin thermostats. We provide tools to generate initial configuration and topology which are based on the elastic network approach and its extensions. Topology generators can be modified by users to add for example a new potential type. The code is written in C/C++ languages and the structure/topology of a molecule is based on an XML format. The code is distributed under GNU public licence and will be available at `http://bionano.icm.edu.pl/`.

1 Introduction

We created an open-source, scalable package for reduced (coarse-grained) molecular dynamics (MD) simulations of biomolecules[1] on a micro- to mili-second time scales. It is written in C/C++ and parallelized with an OpenMP technology. To generate the topology and force field of a molecule we use an XML-based format. Currently implemented force field generators include the Elastic Network Model[2] and its anharmonic extensions for the ribosome[3], nucleosome[4] and HIV-1 protease[5].

2 Molecular Dynamics

MD is a widely used technique to investigate the dynamical properties of molecules. It numerically solves in finite time steps the Newton's equations of motion and provides trajectories i.e., the coordinates and momenta of particles as a function of time. Our MD package generates coarse-grained representations of biopolymers in which entire groups of atoms are represented by single interacting centers (pseudo-atoms), (see Figure 1). In one bead models, one pseudo-atom can represent the whole amino acid or nucleotide. The coarse-graining procedure is applied to reduce the number of degrees of freedom, increase the integration time step, and achieve at least a microsecond MD simulation time scale. RedMD currently supports MD simulations in the microcanonical ensemble, in the canonical ensemble with the Berendsen thermostat, and with the Langevin bath thermal coupling. To enlarge the integration time step we implemented the SHAKE algorithm[6,7] which was developed to satisfy the bond geometry constraints in MD simulations.

Figure 1. The coarse-graining procedure. Left: all-atom representation, right: reduced one-bead model.

2.1 Microcanonical Ensemble

In the microcanonical ensemble (NVE) we consider the Newton's equation:

$$\frac{d\vec{p}_i}{dt} = -\frac{\partial V}{\partial \vec{r}_i} \qquad , \; i = 1, \ldots, N$$

where V is a potential energy function which depends on nucleic coordinates r_i, p_i is the momentum of a particle i and N is the number of particles in the system. Numerical solution of this equation generates the trajectory of motion. We implemented two popular algorithms: velocity Verlet and Leap-Frog[8] which are based on Taylor expansion.

2.2 Thermostats

To maintain constant temperature RedMD provides various thermostats. A common method of both thermal coupling and reproducing the contact with solvent molecules is Langevin Dynamics. We consider the equation:

$$\frac{d\vec{p}_i}{dt} = -\frac{\partial V}{\partial \vec{r}_i} - \gamma_i \vec{p}_i + \vec{R}(t) \qquad , \; i = 1, \ldots, N$$

where γ_i is the collision parameter and $\vec{R}(t)$ is a random force vector satisfying:

$$< \vec{R}(t) >= 0, \;\; < \vec{R}(t)\vec{R}(t') >= 2\gamma_i k_B T m_i \delta(t - t')$$

where k_B is the Boltzmann constant and T is the bath temperature. We solve this equation with Brünger-Brooks-Karplus integrator (BBK)[9]. Through the Langevin equation the system couples to a heat bath globally, but is also locally subjected to random noise. If we are interested in imposing the global coupling with minimal local disturbance the Langevin equation is modified to:

$$\frac{d\vec{p}_i}{dt} = -\frac{\partial V}{\partial \vec{r}_i} + m_i \gamma \left(\frac{T_0}{T} - 1 \right) \vec{p}_i \qquad , \; i = 1, \ldots, N$$

where T_0 is the bath temperature, T the current temperature and γ (in ps^{-1}) determines the strength of coupling with the thermal bath. This is known as the Berendsen thermostat[10]. The solution is based on scaling the momenta in each step by a factor

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)}$$

where τ_T is equal to $(2\gamma)^{-1}$.

2.3 Brownian Dynamics (BD)

RedMD provides a BD simulation based on the Ermak-McCammon algorithm[11]. The trajectory is generated according to the equation:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{D}{k_b T}\vec{F}_i(t)\Delta t + \vec{R}_i(t) \qquad , \; i = 1, \ldots, N$$

where D is a diffusion coefficient of a molecule and $\vec{R}(t)$ is a random force vector satisfying:

$$< \vec{R}(t) >= 0, \;\; < \vec{R}(t)\vec{R}(t') >= 6D\Delta t\delta(t-t').$$

3 Parallelization

Because the calculation of the non-bonded interactions and forces is the most time consuming step in MD, its optimization is very important. To parallelize this calculation we applied the OpenMP technology which uses a shared memory architecture. For the nucleosome model test force calculations for 1000 NVE simulation steps show that the speedup (defined as the ratio of time using one thread to the time of using N threads) for 7 cores is over 6 so it is almost linear.

4 The XML Force Field Format

The initial configuration and force field of the molecule are generated in an XML-based format. We provide utilities to produce the input XML file from the PDB or PDBML/XML files (`http://www.rcsb.org/pdb/`). Our XML format is a flexible way of representing the topology and force field. It contains information on atomic ids, masses, names, coordinates, and properties such as bonded and non-bonded potentials. It is possible to save any calculated values like temperature factors, forces or energies (see example below).

```
<?xml version="1.0"?>
<STRUCTURE>
<!-- Generated with: ./RedMD_genModel_Rib 1A36.txml -->
  <NONBONDED>
    <NBMORSE cutoff="35.000000">
      <PAIR type1="CA" type2="CA" alpha="0.707000" E0="0.055055" l0="9.500000"/>
      <PAIR type1="P" type2="P" alpha="0.707000" E0="0.071348" l0="17.600000"/>
      <PAIR type1="CA" type2="P" alpha="0.707000" E0="0.062674" l0="12.930584"/>
    </NBMORSE>
  </NONBONDED>
  <MOLECULE molId="mol1">
   <GROUP label="basePairs" k="0.600000"/>
   <ATOM id="1" name="P" x="-50.7" y="76.7" z="327.1" resName="U" chainID="A" resSeq="2" m="305"/>
   <ATOM id="2" name="P" x="-50.8" y="73.7" z="332.1" resName="U" chainID="A" resSeq="3" m="305"/>
   <ATOM id="3" name="P" x="-48.5" y="75.0" z="336.9" resName="G" chainID="A" resSeq="4" m="344"/>
   <BOND idAtom1="1" idAtom2="2" k="3.0" l0="5.80"/>
   <BOND idAtom1="1" idAtom2="3" k="2.5" l0="10.15"/>
   <BOND idAtom1="1" idAtom2="4" k="0.5" l0="9.40"/>
   <MORSE idAtom1="1" idAtom2="5" alpha="0.707" l0="6.19" E0="0.71" mark="o:P:P"/>
   <MORSE idAtom1="1" idAtom2="6" alpha="0.707" l0="6.87" E0="0.63" mark="o:P:P"/>
   <MORSE idAtom1="1" idAtom2="7" alpha="0.707" l0="11.90" E0="0.27" mark="o:P:P"/>
 </MOLECULE>
</STRUCTURE>
```

5 From PDB to Trajectory

First, one needs to convert a PDB or PDBML file to our XML file format with one of the programs distributed in the RedMD package. The user can choose from a few coarse-grained force fields. Second, the user needs to specify MD simulation parameters and then

can generate the trajectory. Currently supported trajectory output formats are XYZ, PDB, DCD and VEL (analogous to XYZ but saves velocities).

Acknowledgments

We acknowledge support from University of Warsaw (115/30/E-343/S/2007/ICM BST1255 and 115/30/E-343/BST1345/ICM2008), Polish Ministry of Science and Higher Education (3 T11F 005 30), Fogarty International Center (NIH Research Grant R03 TW07318) and Foundation for Polish Science. Simulations were performed at ICM University of Warsaw grant no G31-4.

References

1. A. Gorecki, M. Szypowski, B. Domagala, and J. Trylska, *RedMD - a coarse-grained molecular dynamics package*, in preparation, 2008.
2. M. Tirion, *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis*, Phys. Rev. Lett., **77**, 1905–1908, 1996.
3. J. Trylska, V. Tozzini, and J. A. McCammon, *Exploring Global Motions and Correlations in the Ribosome*, Biophys. J., **89**, 1455–1463, 2005.
4. K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Langovski, and J. Smith, *Coarse-grained Force Field for the Nucleosome from Self-consistent Multiscaling*, J. Comp. Chem., **in press**, 2008.
5. V. Tozzini, J. Trylska, C. E. Chang, and J. A. McCammon, *Flap Opening Dynamics in HIV-1 Protease Explored with a Coarse-Grained Model*, J. Struct. Biol., **157**, 606–615, 2007.
6. Jean-Paul Ryckaert, Giovanni Ciccotti, , and Herman J.C. Berendsen, *Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkenes*, Journal of Computational Physics, **23**, 327–341, 1977.
7. Timothy R. Forester and William Smith, *SHAKE, Rattle, and Roll: Effcient Constraint Algorithms for Linked Rigid Bodies*, Journal of Computational Chemistry, **19**, 102–111, 1998.
8. Andrew R. Leach, *Molecular Modelling. Principles and Applications*, vol. second edition, Pearson Education Ltd., England, 2001.
9. Jesus A. Izaguirre, Daniel P. Castarello, Justin M. Wozniak, and Robert D. Skeel, *Langevin stabilization of molecular dynamics*, J. chem. Phys., **114**, 2090–2098, 2001.
10. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, *Molecular dynamics with coupling to an external bath*, J. chem. Phys., **81**, 3684–3690, 1984.
11. D.L. Ermak and J.A. McCammon, *Brownian dynamics with hydrodynamics interactions*, J. Chem. Phys., **69**, 1352–1360, 1978.

Insights into the Self-Assembly of Phenylalanine Oligopeptides by Replica Exchange MD Simulations with the GBSW Implicit-Solvent Model

Phanourios Tamamis[1], **Lihi Adler-Abramovich**[2], **Ehud Gazit**[2], **and Georgios Archontis**[1]

[1] Department of Physics, University of Cyprus, PO20537, CY1678 Nicosia, Cyprus
E-mail: tamamis@ucy.ac.cy

[2] Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel

The diphenylalanine peptide (FF), the core recognition motif of the Alzheimer's β-amyloid peptide, self-assembles into tubular structures of high stability. We have studied the aggregation properties of FF and the related triphenylalanine peptide (FFF) by 0.4-μs implicit- solvent Replica Exchange MD simulations of aqueous FF and FFF solutions. The FF and FFF peptides form ellipsoidal aggregates with a similar density and shape in the simulations. Within each aggregate, we observe structural features, which are consistent with the properties of L-Phe-L-Phe crystals. In particular, the aromatic planes of interacting sidechains are mainly oriented perpendicular to each other and the backbone moieties of several (2-6) adjacent peptides interact frequently by head (NH_3^+) -to-tail (^-OOC) hydrogen bonds, forming open or closed (ring-like) linear networks. The ring networks of six peptides observed in the FF simulations are reminiscent of the hexagonal FF rings in the L-Phe-L-Phe crystals. The rings are energetically more stable than the open networks, due to both non-polar and polar interactions. The network propensity is higher in the FFF solution, mainly due to stronger non-polar and to a smaller extent due to stronger polar interactions in the networks of the FFF aggregate; in line with this observation is the somewhat higher stability of the FFF aggregate, observed in the simulations.

1 Introduction

The ability of short peptide fragments to self-assemble into amyloids[1], nanotubes[2] and systems responsive to external stimuli (pH, temperature, concentration of specific solutes)[3] is the focus of intense experimental and computational studies in recent years, as it can provide insights on the formation of amyloid fibers and has potential applications in biomaterial synthesis, nanodevice fabrication and tissue engineering.

The diphenylalanine peptide (NH_2-L-Phe-L-Phe-COOH, FF), the core recognition motif of the Alzheimer's β-amyloid peptide was already been crystallized[4]. Under certain conditions the diphenylalanine peptide self assembles into nanotubes of remarkable stiffness[5], which can serve as casts for the fabrication of silver nanowires[6]. In the crystals, the FF peptides are hydrogen-bonded head-to-tail, forming helical chains with six peptides per helical turn and a 10-Å van der Waals diameter[4]. Adjacent helices are oriented parallel to each other and interact extensively via an intricate three-dimensional stacking arrangement of the aromatic side chains. This structural information provides some hints for the molecular organization of the peptide nanotubes[7]. Nevertheless, the understanding of the key factors responsible for the nanotube stabilization is still not complete. For example, the peptide Ac-Phe-Phe-NH_2 forms highly-ordered tubular structures despite the

lack of charge in its terminal ends, suggesting that the interactions between the aromatic sidechains rather than the electrostatic backbone interactions play the key role in the self-assembly process. At the same time, chemical modifications of the FF terminal ends cause the formation of macroscopic hydrogels[8] or amyloid-like fibers[9], suggesting that the nanostructures formed by FF depend also on the chemical nature and interactions of its terminal ends.

To obtain further insights on the aggregation properties of these systems, in the present study we investigate by MD simulations the properties of aqueous solutions formed by FF and the related system, triphenylalanine peptide (NH_2-FFF-COOH).

2 Systems and Methods

All simulations were performed with the CHARMM program, version c35a1[10]. We simulated two aqueous solutions, consisting of 12 FF dipeptides and 8 FFF tripeptides, respectively. The peptides were placed in a 57-Å cubic box, modeling 34 mgr/ml (FF) and 33 mgr/ml (FFF) solutions. The box was replicated by periodic boundary conditions. The peptide atomic charges, van der Waals and stereochemical parameters were taken from the CHARMM27 all-atom force field[11,12]. The aqueous solvent effects were modeled implicitly by the Generalized Born approximation GBSW[13,14]. To improve the conformational sampling, each solution was simulated by the replica-exchange scheme, with 10 replicas spanning the temperature ranges 289-405 K (FF) and 288-416 K (FFF). The replica temperatures were optimized iteratively in the beginning of the simulations as in ref. 15, 16, targeting a uniform exchange probability of 18-20% among adjacent replicas. The simulation length for each temperature was 40 ns, yielding a total simulation time of 0.4 μs for the 10 replicas. Replica exchanges were attempted at 10-ps intervals. The analysis was done with the CHARMM modules and in-house FORTRAN programs.

3 Results and Discussion

Geometrical Analysis: In both solutions the peptides form approximately ellipsoidal aggregates, which are stable at 300 K. The geometrical properties of the aggregates in the 300-K simulations are summarized in table 1.

Aggregate Properties	FF	FFF
Radius of gyration (Å)	9.48±0.41	9.25±0.30
Volume (Å^3)	3495.77±14.53	3422.04±13.28
Density (gr/ml)	1.779±0.007	1.782±0.006
I1/I2, I1/I3, I2/I3	0.76±0.12, 0.66±0.10, 0.88±0.06	0.76±0.11, 0.66±0.09, 0.86±0.06
PSA (%)	29.00±6.04 (42.18±1.44)	24.40±4.48 (34.64±3.41)

I1, I2 and I3 are the principal moments of inertia of the aggregate. PSA is the polar accessible surface area. The numbers in parentheses correspond to the average PSA of FF and FFF monomers, computed by independent monomer simulations[17].

Table 1. Geometrical characteristics of FF and FFF aggregates at 300 K.

The FF and FFF aggregates have similar shapes (indicated by the comparable moment-of-inertia ratios) and densities. The radius of gyration of the FFF aggregate has a somewhat smaller mean and standard deviation (sd), reflecting the fact that the FFF aggregate is somewhat more stable. In line with this observation, the radius of gyration of the FF aggregate increases faster with temperature[17]. The solvent-accessible surface area of the FF and FFF peptides has a mixed character, due to the non-polar sidechains and the polar (charged) terminal ends and interior peptide bonds. The average fraction of the FF and FFF peptide polar solvent accessible surface area (PSA) is 42.2% and 34.6%, respectively, at 300 K, as computed by simulations of the FF and FFF monomers (with the same implicit model GBSW[13,14]). The PSA ratio is reduced in the FF and FFF aggregates, reflecting the fact that the polar groups participate in several interactions. Indeed, the peptides form network structures in the aggregates, as we analyze below.

Network Structures: The backbone moieties of several (2-6) adjacent peptides are frequently arranged into open or closed (ring-like) linear networks, in which adjacent peptides interact by head (NH_3^+)-to-tail (^-OOC) hydrogen bonds and the aromatic planes of interacting sidechains are mainly oriented perpendicular to each other. A typical, six-peptides ring of the FF simulations is shown in fig. 1, along with the hexagonal ring pattern observed in the L-Phe-L-Phe crystals[4]. The open networks are more frequent due to entropic reasons; nevertheless, the closed networks are more stable energetically, due to both non-polar and polar interactions[17]. The network propensity is higher in the FFF solution. Energetic analysis shows that the non-polar and electrostatic interactions are stronger in the networks of the FFF aggregate[17]; this is in accord with the higher network propensity and the higher stability of the FFF aggregate. The frequency of the peptide networks decreases with temperature, in agreement with the increase in the radius of gyration and the loss of stability[17].

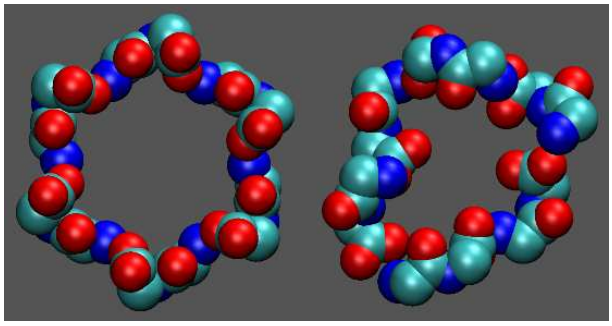


Figure 1. Left: The hexagonal ring network observed in FF crystals[4]. Right: A typical, six-peptides ring network, observed in the FF simulations.

4 Concluding Remarks

We have investigated the conformational properties of aqueous FF and FFF solutions by implicit-solvent Replica Exchange MD simulations. In both solutions, the peptides form stable ellipsoidal aggregates. Within the aggregates the peptides are arranged into open and closed linear networks, which are stabilized by head-to-tail and sidechain interactions and have some of the structural features observed in the FF crystals[4]. The intermolecular interactions are stronger in the FFF system, in line with the higher frequency of inter-peptide networks and the increased stability of the FFF aggregate in the temperature range of the simulations. An energetic analysis shows that the side chains contribute to the aggregate stability by forming direct interactions and by modulating the screening of the termini electrostatic interactions by solvent.

References

1. S. Gilead and E. Gazit, Supramolecular Chemistry 17, 2005, 87-92.
2. E. Gazit, Chem. Soc. Rev. 36, 2007, 1263-1269.
3. K. Chockalingam, M. Blenner and S. Banta. Protein Engineering, Design & Selection 20, 2007, 155-161.
4. C. H. Gorbitz, Chem. Commun. 2006, 2332-2334.
5. N. Kol, L. Adler-Abramovich, D. Barlam, R. Z. Shneck, E. Gazit and I. Rousso. Nano Lett. 5, 2005, 1343–1346.
6. M. Reches and E. Gazit. Science 300, 2003, 625-628.
7. M. Reches and E. Gazit, Nat. Nanotechnol. 1, 2006, 195-200.
8. A. Mahler, M. Reches, M. Rechter, S. Cohen and E. Gazit. Adv. Mater. 18, 2006, 1365-1370.
9. M. Reches and E. Gazit. Isr. J. Chem. 45, 2005, 363-371.
10. B.R. Brooks et al. J. Comput. Chem. 4, 1983, 187-217.
11. N. Foloppe, A. D. MacKerell Jr. J. Comp. Chem. 21, 2000, 86-104.
12. N. Foloppe, A. D. MacKerell Jr. J. Comp. Chem. 21, 2000, 105-120.
13. J. Chen, W. Im, and C. L. Brooks III. J. Am. Chem.Soc. 128, 2006, 3728-3736.
14. W. Im, J. Chen, and C. L. Brooks III. Adv. Prot. Chem. 2006, 72, 171-195.
15. K. Hukushima and K. Nemoto. J. Phys. Soc. Jpn. 65, 1996, 1604-1608.
16. K.Y. Sanbonmatsu and A.E. Garcia. Proteins 2, 2002, 46, 225-234.
17. P. Tamamis, G. Archontis. L. Adler-Abramovich and E. Gazit. Paper in preparation.

Analysis and Classification of the Structural Interactome

Joan Teyra, Maciej Paszkowski-Rogacz, Gerd Anders, and M. Teresa Pisabarro

BIOTEC TU Dresden, Tatzberg 47/49, 01307 Dresden, Germany
E-mail: joan.teyra@biotec.tu-dresden.de

1 Introduction

Protein interactions are essential for intra-cellular communication in biological processes. Proteins are composed of small units or domains that can physically interact together forming multi-domain protein complexes. A single protein can have several binding regions, and each region can engage distinct ligands, either simultaneously or at successive stages of signalling. Detailed information about protein interactions is critical for our understanding of the principles governing protein recognition mechanisms. The structures of many proteins have been experimentally determined in complex with different ligands bound either in the same or different binding regions. Thus, the structural interactome requires the development of tools to classify protein binding regions. A proper classification may provide a general view of the regions that a protein uses to bind others and also facilitate a detailed comparative analysis of the interacting information for specific protein binding regions at atomic level. Such classification might be of potential use for deciphering protein interaction networks, understanding protein function, rational engineering and design. We present the SCOWLP database and web-interface [1, 2], a framework to study protein interfaces and for comparative analysis of protein family binding regions (PBRs).

2 Methodology

SCOWLP was developed following several steps:

2.1 Extraction of Interfaces and Contacting Domains

An accurate definition of the interacting residues is crucial to have a proper clustering of a family PBR. Our database includes all protein-interacting components of the PDB including peptides and solvent, which until now have been excluded from systematic protein interface analysis and databases. The inclusion of water enriches the definition of protein interfaces by considering residues interacting exclusively by water, defined as wet spots [3]. In our database all interface interactions are described at atom, residue and domain level by using interacting rules based on atomic physicochemical criteria [1]. The definition of a domain was extracted from the SCOP database [4]. We consider ”interface” all domain-domain interactions ; that means those belonging to the same protein and also to different proteins. SCOWLP contains 79,803 interfaces contained in 2,561 SCOP families. We grouped the domains participating in each interface by SCOP families, obtaining for each family a list of contacting domains with the residues forming part of the binding region.

2.2 Pair-Wise Structural Alignments (PSAs)

A reliable alignment is indispensable to calculate the similarities among binding regions. For this purpose we used MAMMOTH, which has shown proven accuracy to structurally align protein families [5]. We performed all-against-all PSAs of the contacting domains for each family to be able to measure the similarity among binding regions. SCOWLP contains about 160,000 contacting domains uneven distributed by families. This represents 276 million PSAs performed in a cluster of five Pentium IV 2.6 GHz. The alignments were performed taking the C atoms into account and using a gap penalty function for opening and extension [6]. The root-mean-squared deviation (RMSD) was not considered for measuring the similarity between two interfaces, as the superimposed members of the same family share a common structure.

2.3 Similarity Index (Si)

The residues described in SCOWLP forming and interface were mapped onto the domain-pair structural alignment. We calculated a similarity index (Si) based on the number of interacting residues that overlap and the length of both interacting regions by:

$S_i(a,b) = \frac{2IR_{overlap}(a,b)}{IR_{length(a)} - IR_{gaps(a)} + IR_{length(b)} - IR_{gaps(b)}}$, where a and b represent the two domain structures aligned. The number of interacting residues that match in the PSA is represented by IRoverlap(a, b). This value is divided by the average number of the interacting residues in both domains excluding the interacting residues located in gap regions in the structural alignment (IRgaps).

2.4 Clustering Binding Regions

Based on the calculated Si, we clustered the binding regions of each SCOP family using the agglomerative hierarchical algorithm following several steps: 1) Define as a cluster each contacting domain. 2) Find the closest pair of clusters and merge them into a single cluster. 3) Re-compute the distances between the new cluster and each of the remaining clusters. 4) Repeat steps 2 and 3 until all contacting domains are clustered into a single cluster. To re-compute the distances we used the complete-linkage method, which considers the distance between two clusters to be equal to the minimum similarityof the two members.

2.5 Binding Region Definition by Si Cut-Offs

The result of the clustering can be represented in an intuitive tree or dendrogram, which shows how the individual contacting domains are successively merged at greater distances into larger and fewer clusters. The final PBRs depend on the Si cut-off that is set up. Based on our observations of a representative group of families we set up an empirical maximum similarity cut-off value of 0.4. We pre-calculated the results for Si cut-offs at 0, 0.1, 0.2, 0.3 and 0.4 to offer a range of values that allow flexibility in the final analysis of PBRs. The SCOWLP web application offers the possibility to display the classification at any of these cut-off values. Our classification clustered 160,000 contacting domains from 2,561 families in 9,334 binding regions. About 65% of the families contain more than one binding region. These values are obtained for similarity zero and may vary depending on the similarity cut-off applied.

2.6 Interface Definitions

In order to differentiate binding regions having single-interfaces from multi-interfaces, we identified in each binding region the partner for each contacting domain. Each binding region was divided in sub-clusters when there were different domain families interacting in the same binding region. This resulted in a total of 10,300 interfaces. The classification shows a 78% of the binding regions having a single-interface and the rest having mainly 2 or 3 interfaces per region. These numbers have to be carefully interpreted by taking into account the limitation of the structural information contained in the PDB (i.e. 1,715 binding regions contain a unique member in the PDB and therefore only one known interface per binding region).

3 Conclusions

SCOWLP database contains detailed interacting information of protein interfaces and the hierarchical classification of PBRs. It represents a framework to study protein interfaces and for comparative analysis of protein family binding regions. This comparison can be performed at atomic level and allows the user to study interactome conservation and variability. The new SCOWLP classification may be of great utility for reconstruction of protein complexes, understanding protein networks and ligand design. The web application is available at `http://www.scowlp.org`.

References

1. J. Teyra, A. Doms, M. Schroeder, M. T. Pisabarro, *SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces*, BMC Bioinformatics. **7**, 1004, 2006.
2. J. Teyra, M. Paszkowski-Rogacz, G. Anders, M. T. Pisabarro, *SCOWLP classification: structural comparison and analysis of protein binding regions*, BMC Bioinformatics. **9**, 9, 2008.
3. J. Teyra, M. T. Pisabarro, *Characterization of interfacial solvent in protein complexes and contribution of wet spots to the interface description*, Proteins **67**, 2577–2587, 2007.
4. `http://scop.mrc-lmb.cam.ac.uk`.
5. D. Lupyan D, A. Leo-Macias A, A. R. Ortiz, *A new progressive-iterative algorithm for multiple structure alignment*, Bioinformatics **21**, 3255–3263, 2005.
6. A. R. Ortiz, C. E. Strauss, O. C. Olmea, *MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison*, Protein Sci **11**, 2606–2621, 2002.

Characterization of the Binding Surface of the Human Protein GABARAP

Yvonne Thielmann[1], Oliver H. Weiergräber[1], Jeannine Mohrlüder[1,2], Bernd W. König[1,2], Rudolf Hartmann[1], Thomas Stangler[1,2], Katja Wiesehan[1], and Dieter Willbold[1,2]

[1] Institut für Neurowissenschaften und Biophysik, Research Centre Jülich, 52425 Jülich, Germany
E-mail: {y.thielmann, d.willbold}@fz-juelich.de

[2] Institut für Physikalische Biologie und BMFZ, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

GABA(A) receptors are ligand-gated chloride channels that mediate inhibitory neurotransmission. The $GABA_A$ receptor associated protein (GABARAP) interacts with the gamma2 subunit of the $GABA_A$ receptor, modulates channel kinetics and promotes receptor clustering. Two hydrophobic pockets acting as indole binding sites were identified as major determinants of the ligand specificity of GABARAP by two dimensional NMR. We identified peptide K1 that binds GABARAP with high affinity. Co-crystals of GABARAP and K1 diffract to 1.3 Å resolution. Each hydrophobic pocket of GABARAP is occupied by a tryptophan residue of the peptide. Recently we found that calreticulin binds GABARAP. Co-crystals of GABARAP and calreticulin (178-188) diffract to a resolution of 2.3 Å. In this case the two hydrophobic pockets are occupied by a tryptophan and a leucine, respectively. This is the first complex structure of GABARAP with a native ligand.

1 GABARAP Displays Two Hydrophobic Pockets

The role of tryptophan as a key residue for ligand binding to the ubiquitin-like modifier GABARAP was investigated. Two tryptophan binding hydrophobic patches were identified on the conserved face of the GABARAP structure by NMR spectroscopy and molecular docking (Figure 1). GABARAP binding of indole and indole derivatives including the free amino acid tryptophan was quantified. The two tryptophan binding sites can be clearly distinguished by mapping the NMR-derived residue specific apparent dissociation constant, K_d, onto the three-dimensional structure of GABARAP. The biological relevance of tryptophan binding pockets of GABARAP is supported by a highly conserved tryptophan residue in the GABARAP binding region of calreticulin, clathrin heavy chain, and the gamma2 subunit of the $GABA_A$ receptor. Replacement of tryptophan by alanine abolishes ligand binding to GABARAP[1].

2 Co-Crystallization of GABARAP with Ligand Peptides

Next we have determined the X-ray structure of the soluble form of human GABARAP complexed with a high-affinity synthetic peptide at 1.3 Å resolution (Figure 2). The data shed light on the probable binding modes of key interaction partners, including the $GABA_A$ receptor and the cysteine protease Atg4. The resulting models provide a structural background for further investigation of the unique biological properties of GABARAP[2].

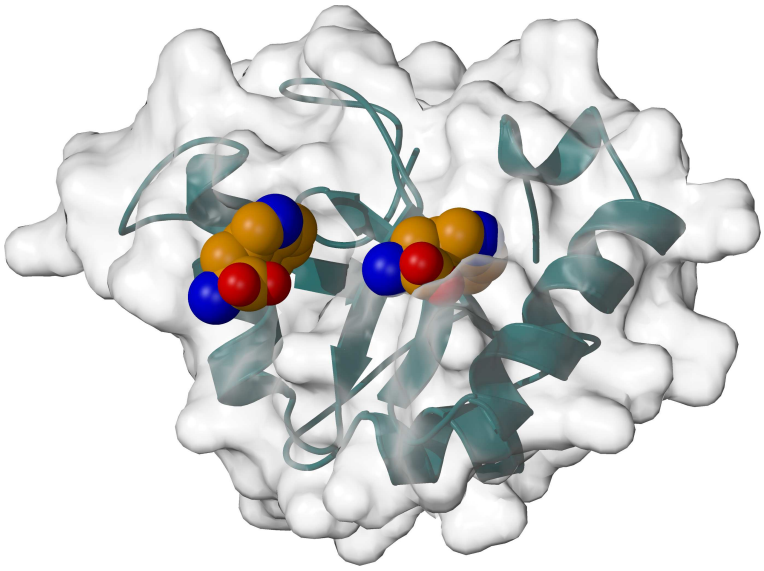


Figure 1. GABARAP in ribbon and surface representation. Two indole molecules are placed in the hydrophobic pockets HP1 and HP2, respectively.

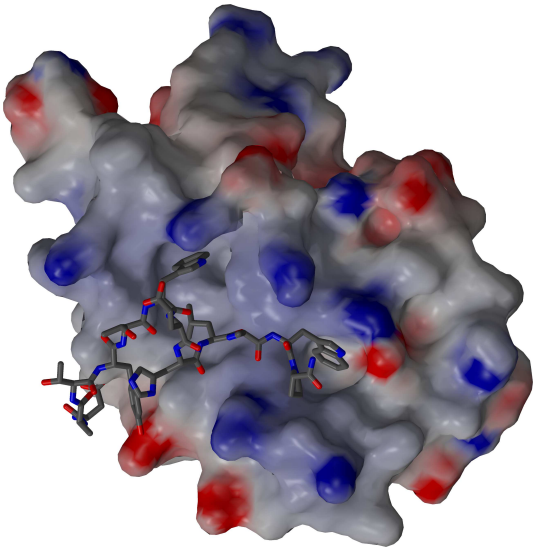


Figure 2. Overview of the GABARAP-K1 complex. GABARAP is depicted as a surface representation. The peptide is shown in stick mode in dark gray. Both tryptophans are deeply buried in the hydrophobic pockets.

Recently we have succeeded in co-crystallizing GABARAP and a fragment of calreticulin and have determined the X-ray structure at 2.3 Å resolution (Figure 3). The results

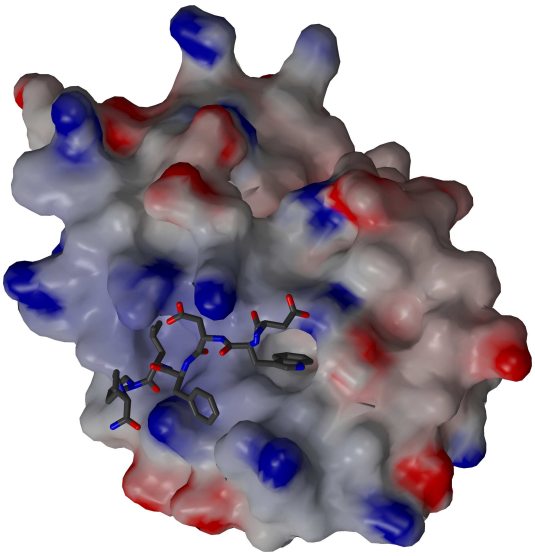


Figure 3. Overview of the GABARAP-CRT(178-188) complex. GABARAP is depicted as a surface representation. The peptide is shown in stick mode in dark gray. The tryptophan and leucine residues are buried in HP1 and HP2, respectively.

improve our understanding of the GABARAP-calreticulin interaction and serve as a model for the interaction with clathrin heavy chain and NSF. The derived model of the complex with full length calreticulin provides a starting point for the investigation of multimeric complexes with additional binding partners of GABARAP.

Acknowledgments

This study was supported by a research grant from the Deutsche Forschungsgemeinschaft (DFG) to D.W. (Wi1472/5).

References

1. Thielmann, Y., Mohrlüder, J., Koenig, B. W., Stangler, T., Hartmann, R., Becker, K., Höltje, H.-D. & Willbold, D. (2008). An indole binding site is a major determinant of the ligand specificity of the GABA type A receptor-associated protein GABARAP. Chem. Biochem., in press.
2. Weiergräber O. H., Stangler T., Thielmann Y., Mohrlüder J., Wiesehan K. & Willbold D. (2008). Ligand binding mode of $GABA_A$ receptor-associated protein. Submitted.

Freezing and Collapse of Flexible Polymers

Thomas Vogel, Michael Bachmann, and Wolfhard Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ), Universität Leipzig,
Postfach 100 920, D-04009 Leipzig, Germany
E-mail: {*vogel, bachmann, janke*}*@itp.uni-leipzig.de*

We analyze the freezing and collapse transition of a simple model for flexible polymer chains on simple cubic and face-centered cubic lattices by means of sophisticated chain-growth methods. In contrast to bond-fluctuation polymer models in certain parameter ranges, where these two conformational transitions were found to merge in the thermodynamic limit, we conclude from our results that the two transitions remain well-separated in the limit of infinite chain lengths. The reason for this qualitatively distinct behavior is presumably due to the ultrashort attractive interaction range in the lattice models considered here.

1 Introduction

It is well known that single homopolymer chains undergo a structural coil–globule transition at the so-called Θ-point. Much theoretical, experimental, and algorithmic work was, and still is, spent to localize that point for various homopolymer models. In principle, there is no longer any difficulty to investigate lattice models up to very long chain lengths. A well working technique to deal with the problem is the (n)PERM algorithm.[1,2]

Since a relatively short time, thanks to generalized-ensemble versions of PERM,[3,4] it is also possible to investigate the temperature range far below the Θ-temperature, where another transition called ground-state–globule, liquid–solid (crystallization) or freezing transition can occur.[3,5] A recent study of the bond-fluctuation model with respect to these different transitions for example showed, that the crystallization and the coil-globule transitions may, but generally do not, coincide in the thermodynamic limit, depending on the interaction range.[5]

In this work, motivated by above mentioned studies, we will report results for the structural transitions of Interacting Self-Avoiding Walks (ISAW) on the simple cubic (sc) and face-centered cubic (fcc) lattice, the simplest model for flexible, interacting polymers.[6]

2 Model and Methods

In the ISAW model, a local attractive interaction between non-bonded nearest neighbors is assumed. The total energy of an ISAW is given by $E = -n$, where n is the number of such contacts. To simulate the model, we use the nPERM algorithm for the study of the Θ-transition of very long chains and a generalized-ensemble version for the simulation of the low-temperature behavior of short polymers. The PERM method is a chain-growth algorithm based on the Rosenbluth method.[7] It includes population control by pruning or enriching populations during the growth, depending on threshold weights.[1–4]

3 Results and Discussion

For investigating structural transitions of the polymer model, we calculate the specific heat and analyze its peak structure. It is expected that even for polymers of finite length, peaks of fluctuating quantities signalize conformational activity.

Figure 1 (left) shows specific-heat peaks of polymers on the sc lattice with lengths $8 \leq N \leq 125$. We see that there is no uniform scaling behavior of the peaks as it was found for the bond-fluctuation model.[8] A first view does not show any regularity at all regarding the low-temperature peaks. However, the dependence of the peak temperatures on the chain length exhibits more systematics, see Fig. 1 (right).

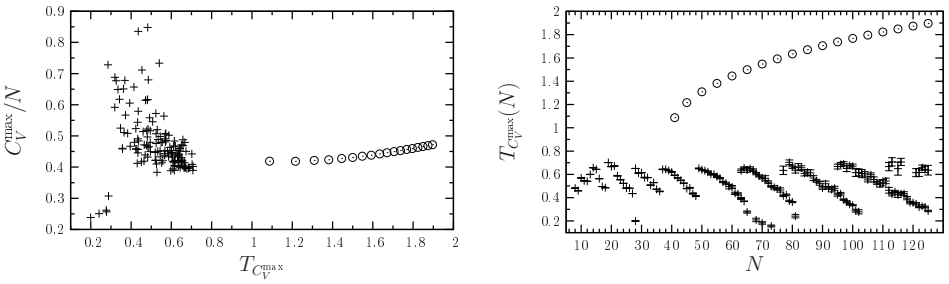


Figure 1. Left: Map of specific-heat maxima for several chain lengths. Circles ($\odot$) symbolize the peaks (if any) identified as signals of the collapse ($T_{C_V^{\max}} > 1$). The low-temperature peaks ($+$) belong to the excitation/freezing transitions ($T_{C_V^{\max}} < 0.8$). Right: Collapse and excitation/freezing peak temperatures of the same specific-heat peaks.

The freezing-transition temperatures show a sawtooth-like behavior which is due to optimal monomer alignment to the underlying lattice. At the lowest peak temperatures, we find chains with very compact ground states which are arranged as cubes or compact cuboids, respectively, e.g., for $N = 27, 36, 48$, etc. The corresponding chains have an energy gap of $\Delta E = 2$ between the ground state and the first excited state. For the sc lattice, this can easily be explained. The first excited state can be constructed by removing a monomer from the corner of the compact state (breaking 3 contacts) and placing it somewhere at the surface (gaining 1 contact). All these chains have a very pronounced freezing-transition peak (not shown). On the other hand, at the other side of the "teeth", the respective chains with one more monomer reside. Here, the formerly pronounced low-temperature peak becomes very weak.

Let us briefly say some words on the Θ-point. Despite of very precise measurements, the nature of this transition is not yet completely understood, considering for example predicted logarithmic corrections[9] which could not be resolved so far in numerical data. Figure 2 shows an illustrating picture of the studied objects (left) and data of transition temperatures for different chain lengths as well as fits to the data (right). There are several approaches to extrapolate T_Θ by fitting $T_c(N)$. We used here the mean-field like fit[6]

$$\frac{1}{T_c(N)} - \frac{1}{T_\Theta} = \frac{a_1}{\sqrt{N}} + \frac{a_2}{N}, \tag{1}$$

which was found to be consistent with numerical data obtained in grandcanonical analyses of lattice homopolymers and the bond-fluctuation model.[10]

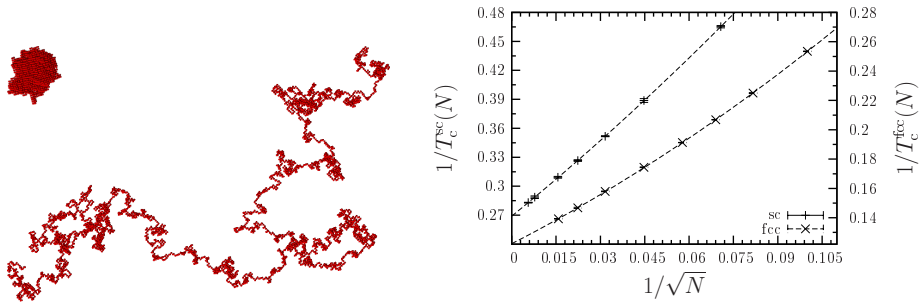


Figure 2. Left: Two typical conformations of a 4096mer on a sc lattice at different temperatures $T \ll T_\Theta$; $E = -5367$ and $T \gg T_\Theta$; $E = -1054$. Right: Inverse collapse temperatures for several chain lengths on sc ($N \leq 32\,000$) and fcc lattices ($N \leq 4\,000$). Drawn lines are fits according to Eq. (1). Error bars are shown, but may not be visible.

Optimal fit parameters using the data in the intervals $200 \leq N \leq 32\,000$ (sc) and $100 \leq N \leq 4\,000$ (fcc) were found to be $T_\Theta^{sc} = 3.72(1)$, $a_1 \approx 2.5$, and $a_2 \approx 8.0$ (sc), and $T_\Theta^{fcc} = 8.18(2)$, $a_1 \approx 1.0$, and $a_2 \approx 5.5$ (fcc). This agrees very well with data published so far,[1,11] but does unfortunately not solve the problem of correction terms either. A very detailed analysis including different fit ansätze and parameter estimates can be found in Ref. 6.

To summarize, there exists a clear low-temperature freezing transition below the Θ-point, which is strongly influenced by lattice restrictions. It is shifted with increasing chain length to lower temperatures ($T_{C_V^{max}}(N) \approx 0.4$) and jumps at chains with very compact ground states back to a value of $T_{C_V^{max}}(N) \approx 0.6$. The temperature interval, in which the freezing peaks fluctuate, does not change when varying the chain length, whereas the finite-length Θ-temperature does.

Acknowledgments

We thank K. Binder, W. Paul, F. Rampf, and T. Strauch for helpful discussions. This work is partially supported by the DFG under Grant No. JA 483/24-1/2, the Graduate School "BuildMoNa", the DFH-UFA PhD College CDFA-02-07 and by a computer-time grant of the NIC at Forschungszentrum Jülich, under No. hlz11.

References

1. P. Grassberger, Phys. Rev. E, **56**, 3682, 1997.
2. H.-P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, J. Chem. Phys., **118**, 444, 2002.
3. M. Bachmann and W. Janke, Phys. Rev. Lett., **91**, 208105, 2003; J. Chem. Phys., **120**, 6779, 2004.
4. T. Prellberg and J. Krawczyk, Phys. Rev. Lett., **109**, 7583, 2004.
5. W. Paul, T. Strauch, F. Rampf, and K. Binder, Phys. Rev. E, **75**, 060801(R), 2007.
6. T. Vogel, M. Bachmann, and W. Janke, Phys. Rev. E, **76**, 061803, 2007.
7. M.N. Rosenbluth and A.W. Rosenbluth, J. Chem. Phys., **23**, 356, 1955.

8. F. Rampf, W. Paul, and K. Binder, Europhys. Lett., **70**, 628, 2005.
9. B. Duplantier, Europhys. Lett., **1**, 491, 1986.
10. N.B. Wilding, M. Müller, and K. Binder, J. Chem. Phys., **105**, 802, 1996.
11. K. Kremer, in: Monte Carlo and Molecular Dynamics of Condensed Matter Systems, K. Binder and G. Ciccotti (Eds.). Editrice Compositori, Bologna, 1996, p. 669.

A Computational Approach to Study the Energy Transduction Mechanism in the Na^+/K^+-ATPase

Christian Weidemüller and Karin Hauser

Institut für Biophysik, Johann Wolfgang Goethe-Universität,
60438 Frankfurt am Main, Germany
E-mail: {hauser,weidemueller}@biophysik.uni-frankfurt.de

The Na^+/K^+-ATPase pumps ions across the membrane which is necessary for maintaining the membrane potential. The energy for this active ion transport is provided by binding and hydrolysis of ATP and has to be transferred from the cytoplasmic nucleotide binding site to the transmembrane domain of ion transport. This transport cycle can also be induced experimentally by applying voltage jumps across the membrane. We simulated the applied electric field by an ionic capacitor and studied the impact on the Na^+/K^+-ATPase by a combination of multiconformation continuum electrostatics (MCCE) and molecular dynamics (MD). Our calculations show a selective activation of the helices M5, M6 and M8 by the electric field. Those helices are likely to act as energy transduction elements.

1 Introduction

The Na^+/K^+-ATPase is an integral membrane ion pump belonging to the superfamily of P-type ATPases. All members of the subfamily (P-type ATPases II) share common structural similarities. These proteins consist of three cytoplasmic domains (**n**ucleotide binding-, **a**ctuator- and **p**hosphorylation domain) and a transmembrane domain (with the ion binding sites) which is composed of 10 α-helices (M1-M10). The cytoplasmic domain is connected to the transmembrane domain by the stalk region (S2-S5). Because of this structural and functional similarity, the mechanism of pumping ions across a membrane is supposed to be common for all P-type ATPases and is described by the Post-Albers scheme (with two main conformational states E1 and E2). In the reaction cycle of the Na^+/K^+-ATPase, three Na^+-ions are pumped out and two K^+-ions are pumped into the cell. Thus the Na^+/K^+-ATPase is electrogenic which is essential for maintaining the membrane potential. Cations have to be transported actively against a gradient. Therefore energy is needed which is provided at the nucleotide binding site by binding and hydrolysis of ATP. This energy has to be transferred to the transmembrane ion binding sites which are located approx. 50 Å apart. Experiments show that the transport cycle can not only be induced by ATP but also by an electric field that is applied to the membrane[1]. These voltage-clamp fluorometry experiments indicate an important role of the highly conserved transmembrane helix M5 and M6 for the energy transduction mechanism of the Na^+/K^+-ATPase. To study the energy transduction theoretically, a combination of electrostatic calculations (MCCE) and molecular dynamic simulations (MD) is used to evaluate the impact of a simulated electric field on the Na^+/K^+-ATPase.

2 Methods

Multiconformation continuum electrostatic (MCCE) calculations simulate simultaneously the residue ionization and side chain rotamers[2,3]. The electric field that is applied to the

transmembrane region of the Na^+/K^+-ATPase (pdb-id: 3B8E ; resolution: 3.5 Å ; E2 enzyme state)[4] was simulated by an "ionic capacitor". Therefore ions were placed above and below the protein and parameterized in MCCE. The atomic model of the Na^+/K^+-ATPase was embedded in a POPC (palmitoyl oleoyl phosphatidylcholine) membrane and the gap between the protein and the membrane was closed with a molecular dynamic (MD) simulation (NAMD)[5]. The exact position of the capacitor and the number of ions that were inserted (strength of the applied electric field) were varied to study the influence of the simulation setup. To include conformational dynamics and protein backbone flexibility, structural snapshots of the MD simulations have been chosen as input coordinates for the electrostatic calculations. The impact of the electric field on the helices M1-M10 and the stalk region S2-S5 was observed by the number of residues per helix that change their conformer. Conformer changes consider both, rotamer changes and changes of the residue ionization.

3 Results and Conclusions

Number of Residues per Helix										
Helix	M1	M2+S2	M3+S3	M4+S4	M5+S5	M6	M7	M8	M9	M10
	21	33	26	36	39	21	23	20	19	21
Number of Conformer Changes for Selected MD Snapshots										
time [ps]	M1	M2+S2	M3+S3	M4+S4	M5+S5	M6	M7	M8	M9	M10
0	2	6	5	8	14	6	3	6	2	3
125	2	10	5	8	11	6	2	8	1	4
250	1	5	3	8	13	6	4	7	1	5
500	1	7	4	12	12	6	2	6	3	5
750	1	6	2	8	11	7	2	7	1	5
1000	3	7	4	8	12	6	2	8	3	5
Averaged Residue Changes per Helix [%]										
	7.9	20.7	14.7	22.2	31.2	29.4	10.9	33.3	8.8	21.4

Table 1. Impact of an electric field on the transmembrane helices of the Na^+/K^+-ATPase including the stalk region.

Table 1 shows that the helices M2, M4, M5, M6 and M8 and M10 respond intensely with residues conformer changes to an applied electric field. The helices M5, M6 and M8 change approximately one-third of their conformers. In addition, the conformer changes of these helices are evenly distributed across the complete helix and not only at the helix ends close to the capacitor as it is the case for the helices M2, M4 and M10. The impact of the simulation setup was tested by varying the geometry of the ionic capacitor and the strength of the applied electric field. Independent of the setup, the number of conformers that change remained high and nearly the same for the helices M5, M6 and M8 as compared to the other transmembrane helices. The strength of the electric field has to be simulated much

higher than the fields that were applied in the experimental setup in order to obtain any effects on conformer occupancy. This can be explained by the limited backbone flexibility in our calculations. However, a qualitative analysis is feasible by this simulation approach. The different structural MD snapshots showed only a small variation in conformer changes as compared to the crystal structure coordinates. In particular, the changes on the helices M5, M6 and M8 remained high for the different structural snapshots. Thus these helices are selectively activated by the electric field supporting the experimental hypothesis that these helices are likely to act as energy transduction elements. The contribution of the helices M5 and M8 to the energy transduction mechanism could also be concluded from our previous electrostatic calculations with modeled structures of the Na^+/K^+-ATPase and with the Ca^{2+}-ATPase, another member of the P-type ATPase family[6]. Further support is provided by calculations on a mutant of the Na^+/K^+-ATPase. This mutant (N776D) shows no voltage dependence in experiments[7] what could be confirmed by our calculations. The number of conformer changes due to the applied electric field is also reduced for M5, M6 and M8. The number of conformer changes is reduced to two-third for the helix M6 and even to the half for the helices M5 and M8.

Acknowledgments

We thank M. Gunner (City College of New York) for providing us with the MCCE code and cluster access at the Frankfurt Center of Scientific Computing (CSC) is acknowledged.

References

1. S. Geibel, J.H. Kaplan, E. Bamberg and T. Friedrich, *Conformational dynamics of the Na^+/K^+-ATPase probed by voltage clamp fluorometry*, Proc. Natl. Acad. Sci. U.S.A. **100**, 964-969, 2003.
2. R. E.Georgescu, E.G. Alexov and M.R. Gunner, *Combining conformational flexibility and continuum electrostatics for calculating pKs in proteins*, Biophys. J. **83**, 1731-1748, 2002.
3. K. Hauser and A. Barth, *Side-chain protonation and mobility in the sarcoplasmic reticulum Ca^{2+}-ATPase: implications for the proton countertransport and Ca^{2+} release* Biophys. J. **93**, 3259-3270, 2007.
4. J. P. Morth, B.P. Pedersen, M.S. Toustrup-Jensen, T.L.M Sorensen, J. Petersen, J.P Andersen, B. Vilsen and P. Nissen, *Crystal structure of the sodium-potassium pump*, Nature **450**, 1043-1049, 2007.
5. J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale and K. Schulten, *Scalable molecular dynamics with NAMD*, J. Comput. Chem. **26**, 1781-1802, 2005.
6. C. Weidemüller and K. Hauser, *Impact of an electric field on P-type ATPases simulated by electrostatic calculations*, Spectroscopy, in press.
7. J.B. Koenderink, S. Geibel, E. Grabsch, J.J.H.H De Pont, E. Bamberg and T. Friedrich, *Electrophysiological analysis of the mutated Na,K-ATPase cation binding pocket*, J Biol Chem **278**, 51213-51222, 2003.

Improved Hydrogen Bond Potentials to Assess *in Silico* Protein Folding

Maarten G. Wolf[1], **Jocelyne Vreede**[2], **Peter G. Bolhuis**[2], **and Simon W. de Leeuw**[1]

[1] Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands
E-mail: m.g.wolf@tudelft.nl

[2] University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Introducing backbone hydrogen bond potentials that alternate between attraction and repulsion, forcing a fast reordering, leads to fast *in silico* folding of small proteins. However thermodynamic properties cannot be extracted from these simulations and they are not suitable for use in generalized ensemble methods. In this paper we address the development of a hydrogen bond potential that is energy conserving and time-reversible, without loss of efficiency.

1 Introduction

Computational modeling at atomic resolution has the potential to reveal properties that are inaccessible to experiments, but its application is amongst others limited by a sampling problem. This is a consequence of the rugged free-energy landscape observed in detailed protein models.

Hydrogen bonds make an important contribution to the free-energy landscape of a protein and are required to compensate the free-energy cost of burying (partially) charged groups.[1] Although there is no net free energy gain upon hydrogen bond reordering, the associated barrier of 2-3 kT[2] contributes strongly to the ruggedness of the free-energy landscape.

In a previous study we showed that the sampling problem can be reduced by facilitating barrier crossings, in particular those associated with hydrogen bond formation/breaking. We introduced hydrogen bond potentials that alternate between attraction and repulsion (AHBP) forcing a fast reordering of the backbone hydrogen bonds which leads to fast in silico folding of small proteins.[3]

However simulations with these alternating hydrogen bond potentials are not time-reversible and they do not conserve energy. Hence thermodynamic properties cannot be extracted from these simulations and they are not suitable for use in generalized ensemble methods. In this paper we address the development of a continuous hydrogen bond potential that is energy conserving and time-reversible, without loss of efficiency.

2 Continuous Hydrogen Bond Potential

There are a number of requirements that must be satisfied in the new hydrogen bond potential. Firstly, in an md simulation the energy needs to be conserved in order to acquire a proper sampling of the free-energy landscape. For the individual hydrogen bond potentials this can easily be satisfied by expressing the distance interaction function I_d and the angle interaction function I_θ as a cubic spline.

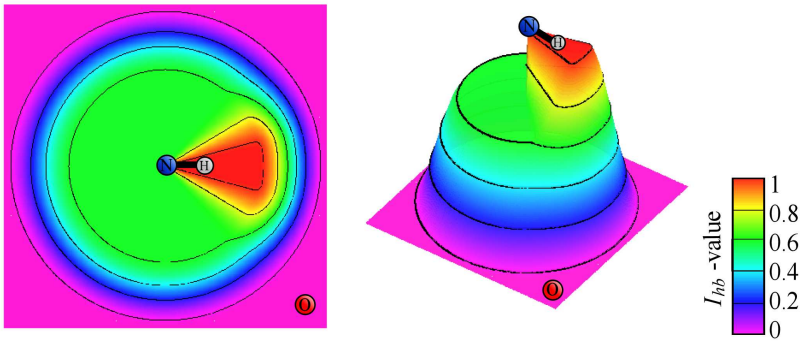


Figure 1. **The interaction function I_{hb} (eq 2) shown for the oxygen position in a plane around the N-H atoms displayed in topview (left) and sideview (right).**

Secondly, the energy and the derived forces of the hydrogen bond potential must result in efficient hydrogen bond reordering. Because reordering mainly requires a change in distance a larger weight for the distance relative to the angle interaction is desired. Within the observation that the angle potential only has meaning at short separation of donor and acceptor the range of the distance and angle function is set to $1.0 - 0.0$ and $1.0 - 0.6$, respectively.

Finally, a hydrogen bond is very selective, usually having only one partner (with occasionally a second), restricting the number of potentials per donor (acceptor) to one. To avoid the need for native state input the partner must be selected based on the conformation in the simulation. Therefore we choose for every donor (acceptor) the partner with the highest value of the hydrogen bond interaction function I_{hb} every time step. In this way the energy is conserved because a switch of partners takes place when the energies are equal. Since the hydrogen bond potential is an interaction function also linear momentum is conserved and only angular momentum is not conserved.

The following continuous hydrogen bond potential V_{hb} (CHBP) is consistent with these requirements.

$$V_{hb}(r) = fc \cdot \sum_{k}^{A,D} \max\{I_{hb}(r_k, r_l)\}_{l \in A,D} \tag{1}$$

$$I_{hb}(r_k, r_l) = I_d(r_k, r_l) \cdot I_\theta(r_k, r_l) \tag{2}$$

$$I_d(r_k, r_l) = \begin{cases} 1 & r_{kl} < r_{min} \\ \frac{2(r_{kl}-r_{min})^3}{(r_{max}-r_{min})^3} + \frac{3(r_{kl}-r_{min})^2}{(r_{max}-r_{min})^2} + 1 & r_{min} \le r_{kl} < r_{max} \\ 0 & r_{max} \le r_{kl} \end{cases} \tag{3}$$

$$I_\theta(\theta_k, \theta_l) = \begin{cases} 1 & \theta_{kl} < \theta_{min} \\ 0.4 \cdot [\frac{2(\theta_{kl}-\theta_{min})^3}{(\theta_{max}-\theta_{min})^3} + \frac{3(\theta_{kl}-\theta_{min})^2}{(\theta_{max}-\theta_{min})^2}] + 1 & \theta_{min} \le \theta_{kl} < \theta_{max} \\ 0.6 & \theta_{max} \le \theta_{kl} \end{cases} \tag{4}$$

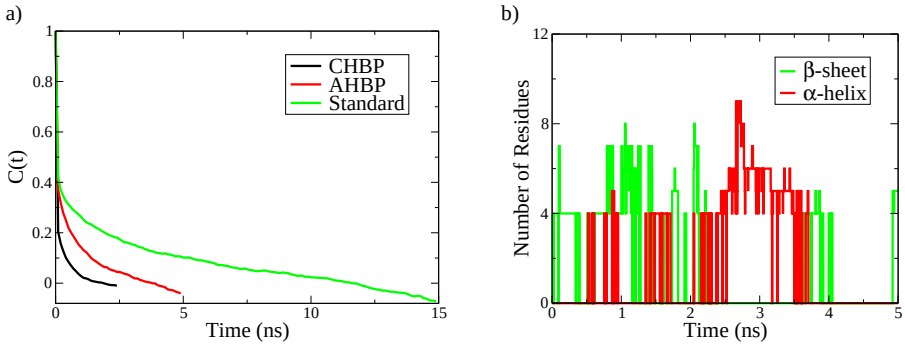


Figure 2. **Effectiveness of the hydrogen bond potentials.** a) The auto correlation function shows the hydrogen bond lifetime for the different simulations. b) The secondary structure formed in a small simulation with the new continuous hydrogen bonds potentials.

with V_{hb} the total potential energy of the hydrogen bond potentials, fc the force constant, r the distance between the donor and acceptor , θ the N-H-O angle and k and l indices of the backbone donor (D) and acceptor (A) atoms.

The parameters for the hydrogen bond potentials were chosen to affect hydrogen bonds that are (very) weak for minimum interference to the system and to leave potentially native hydrogen bonds undisturbed. The minimum and maximum cutoff distances were 0.35 nm and 0.40 nm between the donor and acceptor with a force constant of 50 for the repulsive stage and 0.23 and 0.40 nm between the hydrogen and acceptor with a force constant of -50 for the attractive stage. In both stages the mimimum and maximum N-H-O angle cutoff were 100 and 150.

3 Validation

To validate the new hydrogen bond potential we will show that the requirements 1) energy conservation, 2) selective and 3) effective are satisfied. Therefore we performed NVE simulations with an attractive, with a repulsive and without the hydrogen bond potentials. For these simulations we found a fluctuation of $3.1,\ 4.2$ and $1.8\ kJ\,mol^{-1}$, respectively, around the average total energy $101 \cdot 10^3\ kJ\,mol^{-1}$. This is $\pm 2\%$ of the kinetic and potential energy fluctuation. For all simulations the drift was $-5.7\,kJ\,mol^{-1}\,ps^{-1}$. Within the accuracy of the md algorithm the potentials are thus energy conserving. The energy conservation also shows that a switch of hydrogen bond partners, introducing selectivity, is allowed.

To test the effectiveness of the new hydrogen bond potential we performed NVT simulations where the hydrogen bond potentials are changed every ps in the order attractive-nopotential-repulsive-nopotential-attractive-etc. The effectiveness of these potentials can be expressed as the speed of hydrogen bond reordering, which depends on the hydrogen bond lifetime and the average number of hydrogen bonds. The hydrogen bond lifetime is related to the intramolecular hydrogen bond autocorrelation function, where a faster decrease in the autocorrelation functions indicates a shorter lifetime. From fig. 2a this leads to the following relation between the hydrogen bond lifetime τ_{hb} of the different simulations: $\tau_{hb,CHBP} < \tau_{hb,AHBP} < \tau_{hb,standard}$. The average number of intramolecular

hydrogen bonds is 4.1, 4.7 and 3.9 for the CHBP, AHBP and standard simulation, respectively. These results indicate an even faster hydrogen bond reordering with the new hydrogen bond potentials.

The effectiveness can also be expressed as the speed of forming different secondary structure elements. In fig. 2b the occurrence of α-helix and β-sheet during a simulation with CHBP changing between attrative and repulsive is shown. Clearly both secondary structure elements are easily formed with this potential.

4 Outlook

In this paper we have shown that an energy conserving hydrogen bond potential can result in quick sampling of the conformational space of a polyalanine when the CHBP is changed between attractive and repulsive. However thermodynamic properties can only be extracted from these simulations if this change is Boltzmann weighted. Consequently a method that changes the potentials while maintaining detailed balance, like for example replica exchange, can reveal thermodynamic properties (J. Vreede and M. Wolf, in preparation).

5 Simulation Details

All simulations were performed using the GROMACS[4] software package version 3.3.1 extended with the hydrogen bond potentials using the GROMOS96 43a1 forcefield.[5] A timestep of 2 fs for NVT and 0.1 fs for NVE was used, with all bonds constrained using the LINCS algorithm. Van der Waals interactions were ignored outside a cut-off of 1.2 nm. Electrostatic interactions were treated with the PME method applying a real-space cutoff of 0.9 nm. In the NVT simulation the system temperature was coupled to a Berendsen thermostat.

The starting structure of the simulations is a random collapsed coil poly alanine of 16 residues, dissolved in a box of approximately 2000 SPC water to obtain a density of 0.97 kg/l. Initial velocities were generated randomly to get a system at 300 K, which was simulated under NVT conditions for 1 ns to equilibrate before production runs were performed.

Acknowledgments

This work was funded by the Netherlands Organization for Scientific Research NWO (grant number 635.100.012, program for computational life sciences).

References

1. A. Fernandez, et.al. FEBS Letters **536**, 187-192, 2003.
2. D. van der Spoel, et.al. J. Phys. Chem. B **110**, 4393–4398, 2006.
3. M. G. Wolf and S. W. de Leeuw, Biophys. J. **94**, 3742-3747, 2008.
4. D. van der Spoel, et.al. J. Comp. Chem. **26**, 1701-1718, 2005.
5. W.F. van Gunsteren, et.al. *Biomolecular simulation: the GROMOS96 manual and user guide.* (Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland (1996)).

Sequence Specific DNA Recognition by EcoRV

Mai Zahran[1], **Petra Imhof**[1], **and Jeremy C. Smith**[1,2]

[1] Computational Molecular Biophysics, IWR, University of Heidelberg, Heidelberg, Germany
E-mail: {mai.zahran, petra.imhof}@iwr.uni-heidelberg.de

[2] Center for Molecular Biophysics, Oak Ridge National Laboratory,
P.O. Box 2008 MS6164 Oak Ridge, TN 37831-6164, USA
E-mail: smithjc@ornl.gov

EcoRV is a restriction enzyme produced by Escherichia Coli that destroys invading DNA by cleaving it at a GATATC sequence, as a defence mechanism against viral attacks. EcoRV sharply bends its specific DNA site, by approximately 50°, directly at the center TA step, helping to facilitate correct orientation of the scissile phosphate, the catalytic site of the enzyme and divalent metal ions. We are investigating the physical origins of sequence specificity in EcoRV endonuclease. Computer simulations are performed of three free DNA sequences in aqueous solution, starting from a B-form, in order to understand to what extent the bending is intrinsic to the DNA itself. The results contribute to understanding why the cognate sequence is recognized specifically by the EcoRV restriction enzyme.

1 Introduction

One of the central issues of modern molecular biology and biophysics is to understand the interactions stabilizing complexes in solution, and particularly how a small sequence change can lead to a significant difference in affinity. Protein/DNA interactions provide an important class of biomolecular complexes for studying such issues. Molecular dynamics (MD) simulation techniques, with which individual interactions and dynamics at the atomic level can be probed explicitly, provide a means complementary to experimental techniques, of determining details of molecular structure and interaction.[8] Restriction endonucleases provide interesting model systems for the investigation of sequence-specific protein/DNA interactions. The ability of bacterial cells to resist invading foreign DNA is wholly dependent upon the extraordinarily high fidelity of this recognition process, in which target sites are selected from an enormous molar excess of structurally similar non-specific DNA.[1,3] We are investigating the origins of sequence specificity in EcoRV endonuclease in order to elucidate how molecular interactions and induced fit operate to ensure selectivity for blunt-ended cleavage at the center step of GATATC. EcoRV sharply bends its specific DNA site by approximately 50°, directly at the center TA step, helping to facilitate proper juxtaposition of the scissile phosphate, the catalytic site of the enzyme and divalent metal ions.[7] To examine the origins of EcoRV sequence discrimination, a detailed kinetic and crystallographic study has been performed of the interaction of EcoRV with the cognate sequence GATATC (TA), and two non-cognate sequence, GAATTC (AT) which is the cognate sequence recognized by an other restriction enzyme, EcoRI, and it's mutant, GAAUTC (AU).[2] Examination of DNA binding and bending by equilibrium and stopped-flow fluorescence quenching and fluorescence resonance energy transfer (FRET) methods demonstrate that the capacity of EcoRV to bend the AT sequence site is severely limited, and full bending of AU sequence is achieved at only a threefold reduced rate compared

with the cognate complex (TA). The above results demonstrate that the DNA sequences, although swapping only 2 nucleotides, induce a change of the bending-cleavage mechanism. This change does not imply large conformation adaptations of the protein rather small perturbations leading to the rearrangement of the divalent metal ion binding sites. In the present work, the structures of the free DNA sequences in water are probed. We investigate how the swap of the center base pair leads to a change of the behaviour of the free DNA in water. For this purpose, we performed MD simulations of the three DNA sequences; TA, AT and AU free in aqueous solution.

2 Materials and Methods

Three 14-bp B-DNA molecules were examined with different central nucleotides but the same flanking sequences and same nucleotide content (**TA**-sequence: 5'-dAGAAGA**TA**TCTTGA-3', **AT**sequence: 5'-dAGAAGA**AT**TCTTGA-3', **AU**-sequence: 5'-dAGAAGA**AU**TCTTGA-3'). Standard B-DNA starting structures were generated using the program NAB[6]. The setup of the system was performed with the program CHARMM[9] and the charmm27 force field[10]. Each system was neutralized by adding 28 Na^+ counterions, and an excess of Na^+ and Cl^- ions were added, corresponding to a physiological concentration of 150mN NaCl. The system was solvated with a buffer of explicit water extending of 15 Å in each direction in a cubic box (x=70Å, y=70Å, z=70 Å). Minimizations, heating, equilibration and production runs were performed with the program NAMD[11]. For each of the three systems 20ns MD was performed in a NPT ensemble at 1-atm pressure and 300K. Structural analysis and calculation of the free energy were performed using standard programs; 3DNA[5], Gromacs[4] tools and home made scripts.

3 Results and Discussion

Differences of two nucleotides at the center steps have been shown to be sufficient to hinder full bending and thus cleavage of the DNA in complex with EcoRV restriction endonuclease.[2] We examine here whether or not differences can be seen between these three free DNA molecules in water. We analyzed the molecular origin of bending in terms of local helical parameters of the three DNA structures at the center step calculated with the program 3DNA[5]. The parameters monitored include the local roll angle and the local tilt at the center step. The roll angle measures rotation of a base pair plane about its long axis. This motion creates an angle, narrowing toward the major groove for positive roll, between two otherwise parallel adjacent base pairs. Tilt arises from rotation about a base pair short axis and is generally less than half as large as roll. We calculated the free energy profile at the center step of each DNA sequence as a function of there two parameters. Table 1 shows some local helical parameters for the DNA molecules crystallized in complex with the protein EcoRV, and in the DNA during the MD simulation. The three DNA molecules appear to have different properties for the roll angle at the center step when it is complexed to EcoRV or when it is free in aqueous solution. Indeed the cognate free TA molecule has a minimum free energy for an angle of $10^\circ \pm 0.98^\circ$, whereas the free AT molecule has an angle of $5^\circ \pm 0.65^\circ$ and the free AU molecule is about $3^\circ \pm 0.74^\circ$. Within 5 kJ/mol of the minimum free energy the roll angle for the TA sequence ranges from -10 to +25°,

whereas that for AT reaches only -5 to +15°. The AU sequence explores a wider range than AT but less than the TA sequence, varying between -10 to +20° within 5kJ/mol of the minimum. In contrast to the roll angle, the tilt angle behaves in a similar way for the 3 DNA molecules. The atomic fluctuations (Figure 1) show that the 3 DNA molecules fluctuate in

Parameter	TA[a]	AT[a]	AU[a]	TA[b]	AT[b]	AU[b]
Roll (deg.)	49.9	28.8	53.2	10 ±0.98	5 ±0.65	3 ±0.74
Tilt (deg.)	-3.24	0.87	4.01	0.33 ±0.08	0.63 ±0.05	1.64 ±0.09

[a] DNA crystallographic structures. The respective pdb codes are 1SX8, 2BOD, 2BOE. The values are taken from Hiller[2].
[b] B-DNA structures simulated in water. The values correspond to the angle with the minimum free energy calculated from the 20ns MD simulation.

Table 1. Conformational diffusion constants and corresponding relaxation times.

the same regions, mainly at the 2 termini (atoms 1 to 40; 400 to 500 and 850 to 890 around 0.6 nm for TA sequence, 0.3 nm for AT sequence and 0.8 for AU sequence). The rest of the atoms fluctuate around 0.15 nm, and a periodicity can be observed. The higher values correspond to the fluctuation of the atoms from the backbone (sugar and phosphate), and the lower values to the base pairs. The three structures behave mostly in a very similar

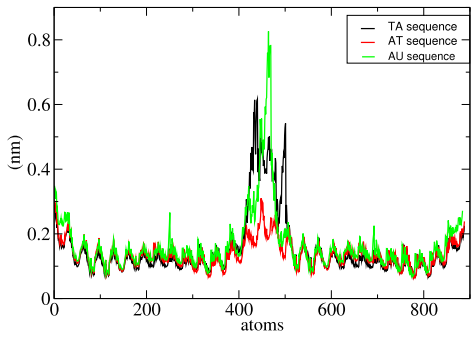


Figure 1. Atomic fluctuation of the 3 DNA sequences during the MD simulation. The first strand is from the atom 1 to 448; the second is from 449 to 891.

way. However the roll angle at the center step shows a none negligible difference.

4 Conclusion

The three DNA molecules simulated here differ by only 1 or 2 nucleotides at the central step. This small difference is sufficient for the EcoRV restriction endonuclease to recognize

only one sequence specifically. The simulations show that the atomic fluctuations and some local helical parameters are very similar for the three sequences, even at the central step. However, the TA sequence appears to be the most flexible of the three, as can be seen by the broader range of roll angles sampled as compared to the other sequences during the 20 ns of the simulations. This finding suggests that recognition of the sequence specifically by the EcoRV restriction enzyme is due in part to inherent structural tendencies of the cognate DNA sequence.

References

1. Kovall, R. A. and Matthews, B. W. *Type II restriction endonucleases: structural, functional and evolutionary relationships*, Curr. Opin. Chem. Biol. **3**, 578-583, 1999.
2. D. A. Hiller, A. M. Rodriguez and J. J. Perona *Non-cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC*, J. Mol. Biol. **354**, 121-136, 2005.
3. D. Parry, S. A. Moon, H. H. Liu,P. Heslop, B. A. Connolly *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures*, J. Mol. Biol. **331**, 1005-1016, 2003.
4. E. Lindahl, B. Hess and D. van der Spoel *GROMACS 3.0: A package for molecular simulation and trajectory analysis*, J. Mol. Mod. **7**, 306-317, 2001.
5. Xiang-Jun Lu and Wilma K. Olson *Non-cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC*, J. Mol. Biol. **31**, 5108-5121, 2003.
6. T. Macke and D.A. Case *Modeling unusual nucleic acid structures. In: Leontes NB, SantaLucia J Jr. , editors.*, In Molecular Modeling of Nucleic Acids.Washington, DC: American Chemical Society, 379-393, 1998.
7. Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, Martin PD, Petratos K, Wilson KS. *The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments.*, EMBO J. **12**, 1781-1795, 1993.
8. Sen S, Nilsson L. *Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation.*, Biophys J. **77**, 1782-1800, 1999.
9. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.*, Journal of Computational Chemistry **4**, 187-217, 1983.
10. MacKerell, Jr. AD, Banavali N, Foloppe N *Development and current status of the CHARMM force field for nucleic acids*, Biopolymers **56**, 257-265, 2001.
11. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. *Scalable molecular dynamics with NAMD.*, Journal of Computational Chemistry **26**, 1781-1802, 2005.

Workshop Program

Monday 19.05.2008

07:40	Pick up at Hotels in Jülich
08:00-09:00	Registration
09:00-09:05	Welcome by Prof. Dr. A. Bachem (CEO, FZ Jülich)
09:05-09:15	Introductory Remarks (Ulrich H.E. Hansmann, NIC-CBB)

Morning Session (Chair: Ulrich H.E. Hansmann)

09:15-09:45	Christoph Pospiech (IBM, Dresden, Germany): Scalable Systems for Computational Biology
09:45-10:30	Volkhard Helms (Saarland University, Saarbrücken, Germany): Computer Simulations of Protein-Protein Association in Water and at Membranes
10:30-11:00	Coffee break
11:00-11:45	Ruben Abagyan (The Scripps Research Institute, La Jolla, USA): Induced Fit in Molecular Docking
11:45-12:30	John Rice (IBM T.J. Watson Research Center, Yorktown Heights, NY, USA): High Performance Computing in Multiscale Modeling Cardiac Contraction: Bridging Proteins to Cells to Whole Heart
12:30-14:00	Lunch

Afternoon Session (Chair: Jan Meinke)

14:00-14:45	Henri Orland (Commissariat à l'Energie Atomique, Gif-sur-Yvette, France): Dominant Pathways in Protein Folding
14:45-15:05	Alexander Schug (University of California San Diego, San Diego, USA): Mutations as Trapdoors: The Rop-dimer with two Competing Native Conformations
15:05-15:25	Nikolay Dokholyan (University of North Carolina at Chapel Hill, Chapel Hill, USA): Simplified Approaches to Complex Biological Systems
15:25-15:45	Mai Suan Li (Polish Academy of Sciences, Warsaw, Poland): New Force Replica Exchange Method and Mechanical Unfolding of Proteins
15:45-16:15	Coffee break

16:15-17:00	Wei Yang (Florida State University, Tallahassee, FL, USA): Advancing Drug and Protein Binding Affinity Predications via Generalized Ensemble Based Methods
17:00-17:20	Iris Antes (Max-Planck-Institut für Informatik, Saarbrücken, Germany): Protein-ligand Docking Including Protein Flexibility – An Hierarchical Approach
17:20-17:40	Slawomir Orlowski (Nicolaus Copernicus University, Torun, Poland): Computer Modeling of Small Ligands Diffusion in Drosophila Melanogaster Hemoglobin
17:40-18:00	Bogdan Lesyng (University of Warsaw, Faculty of Physics, Warsaw, Poland): Protein-ligand Docking with a Two-scale Receptor Dynamics and a QM/MM Interaction Potential
18:00	Welcome Reception and Poster Session
20:30	Bus to Hotels in Jülich

Tuesday 20.05.2008

08:10	Pickup at Hotels in Jülich

Morning Session (Chair: Olav Zimmermann)

09:00-09:45	Andrzej Kolinski (University of Warsaw, Warsaw, Poland): Multiscale Modeling of Protein and Protein Assemblies
09:45-10:05	Shura Hayryan (Academis Sinica, Taipei, Taiwan (ROC)): Some Aspects of RNA Folding Studied by Lattice Simulations
10:05-10:25	Kay Hamacher (TU Darmstadt, Darmstadt, Germany): Coarse-Grained Molecular Models for High-Throughput and Multi-Scale Functional Investigations
10:25-11:00	Coffee break
11:00-11:45	Ron Elber (University of Texas at Austin, Austin, USA): Atomically Detailed Simulations of Kinetics in Molecular Biophysics by Milestoning
11:45-12:05	Anton Feenstra (Free University Amsterdam, Amsterdam, The Netherlands): Predicting Protein Interactions from Functional Specificity using Multi-Relief and Multi-Harmony
12:05-12:25	Sebastian Kmiecik (Selvita, Krakow, Poland): Designing an Automatic Pipeline for Protein Structure Prediction
12:25-12:40	Group-Photo
12:40-14:00	Lunch

Afternoon Session (Chair: Walter Nadler)

14:00-14:45	Wilfred F. van Gunsteren (ETH Zürich, Zürich, Switzerland): Computer Simulation of Biomolecular Systems: Where Do We Stand?
14:45-15:05	Karine Voltz (German Cancer Research Center, Heidelberg, Germany): A Coarse-grained Model for the Nucleosome
15:05-15:25	Maciej Długosz (University of Warsaw, Warsaw, Poland): Interactions of Aminoglycosidic Antibiotics with the 30S Subunit - Brownian Dynamics Study
15:25-15:45	Junalyn Navarra-Madsen (TWU, Denton, USA): Coloring the Mu Transpososome
15:45-16:15	Coffee break
16:15-17:00	Roland Netz (Technical University Munich, München, Germany): Peptide Adhesion and Friction: Theoretical Approaches
17:00-17:20	Rainer Böckmann (Saarland University, Saarbrücken, Germany): Kinetics, Statistics, and Energetics of Lipid Membrane Electroporation Studied by Molecular Dynamics Simulations
17:20-17:40	Borries Demeler (The University of Texas, Health Science Center at San Antonio, San Antonio, Texas, USA): Modeling Conformational and Molecular Weight Heterogeneity with Analytical Ultracentrifugation Experiments (AUC)
17:40-18:00	Wolfgang Fischer (National Yang-Ming University, Taipei, Taiwan): Short Membrane Proteins from Viruses: Channel-pore Dualism?
18:15	Bus to Castle Obbendorf (Hambach) for Dinner
18:30	Dinner (Sponsored by IBM Germany) Greetings by Dr. Sebastian Schmidt, Divisional Director of Research Centre Jülich
22:00	Bus to Hotels in Jülich

Wednesday 21.05.2008

8:10	Pickup at Hotels in Jülich

Morning Session (Chair: Sandipan Mohanty)

09:00-09:45	Michael Feig (Michigan State University, East Lansing, USA): Simulating Biomolecules in Cellular Environments
09:45-10:05	Giovanni La Penna (National Research Council, Sesto Fiorentino, Italy): Modelling the Free Energy of Polypeptides in Different Environments
10:05-10:25	Joachim Dzubiella (Technical University Munich, Garching, Germany): Insights from Atomistic Computer Simulations of Halophilic Proteins
10:25-11:00	Coffee break

11:00-11:45	Philippe Derreumaux (CNRS and University of Paris 7, Paris, France): Simulating the Early Steps of Amyloid Fibril Formation and Disassembly
11:45-12:05	Alfonso De Simone (University of Cambridge, Cambridge, UK): Probing the Prion Hydration by Molecular Dynamics Simulations: From Native via Misfolded to Amyloid Conformations
12:05-12:25	Volker Knecht (Max Planck Institute of Colloids and Interfaces, Potsdam, Germany): Folding and Aggregation of Model Amyloid Peptides in Explicit Solvent and at an Interface
12:25-14:00	Lunch

Afternoon Session (Chair: Ulrich H. E. Hansmann)

14:00-14:20	Horacio Sanchez (Forschungszentrum Karlsruhe, Eggenstein-Leopoldshafen, Germany): High Throughput in-silico Screening against Flexible Protein
14:20-14:40	Michal Wojciechowski (Polish Academy of Sciences, Warsaw, Poland): Effects of Confinement on Protein Folding
14:40-15:25	Dietmar Schomburg (Technische Universität Braunschweig, Braunschweig, Germany): Bioinformatics, Metabolomics, and Systems Biology
15:25-15:30	Concluding Remarks (Ulrich H. E. Hansmann, NIC-CBB)
afterwards	Bus to Cologne
	Social Event: Visit of a Traditional Brewhouse (drinks and food on one's own expense)
21:00	Bus to Jülich

List of Participants

- Abagyan, Ruben; The Scripps Research Institute; La Jolla; USA
- Anand, Priya; Panjab University; Chandigarh; India
- Andrews, Maximilian Nicholas; TU Dortmund; Dortmund; Germany
- Antes, Iris; Max-Planck-Institut für Informatik; Saarbrücken; Germany
- Appel, Heiko; Freie Universität Berlin; Berlin; Germany
- Bahadur, Ranjit Prasad; Jacobs University; Bremen; Germany
- Becker, Caroline; Universität des Saarlandes; Saarbrücken; Germany
- Behringer, Hans; Universität Bielefeld; Bielefeld; Germany
- Berryman, Josh; University of Leeds; Leeds; United Kingdom
- Bischoff, Nils; Universität zu Köln; Köln; Germany
- Biswas, Mithun; Uni Heidelberg; Heidelberg; Germany
- Böckmann, Rainer; Saarland University; Saarbrücken; Germany
- Brunner, Konrad; Universität Regensburg; Regensburg; Germany
- Chebaro, Yassmine; LBT CNRS UPR 9080; Paris; France
- Choutko, Alexandra; ETHZ; Zürich; Switzerland
- Colas, Claire; LBT CNRS UPR 9080; Paris; France
- De Simone, Alfonso; University of Cambridge; Cambridge; United Kingdom
- Decha, Panita; Chulalongkorn University / University of Vienna; Vienna; Austria
- Demeler, Borries; The University of Texas, Health Science Center at San Antonio; San Antonio, Texas; USA
- Deringer, Volker; RWTH Aachen; Aachen; Germany
- Derreumaux, Philippe; CNRS and University of Paris 7; Paris; France
- Despa, Florin; Univ. of Chicago; Chicago; USA
- Dickopf, Thomas; Universität Bonn; Bonn; Germany
- Długosz, Maciej; University of Warsaw; Warsaw; Poland
- Dokholyan, Nikolay; University of North Carolina at Chapel Hill; Chapel Hill; USA

- Dressel, Frank; TU Dresden; Dresden; Germany
- Dzubiella, Joachim; Technical University Munich; Garching; Germany
- Einert, Thomas; Technische Universität München; Garching b. München; Germany
- Elber, Ron; University of Texas at Austin; Austin; USA
- Engin, Özge; Koc University; Istanbul; Turkey
- Feenstra, Anton; Free University Amsterdam; Amsterdam; The Netherlands
- Feig, Michael; Michigan State University; East Lansing; USA
- Fink, Florian; University of Regensburg; Regensburg; Germany
- Fischer, André; Universität Regensburg; Regensburg; Germany
- Fischer, Wolfgang; National Yang-Ming University; Taipei; Taiwan
- Fließ, Anja; Volkswagen-Stiftung; Hannover; Germany
- Fulle, Simone; J.W. Goethe-University; Frankfurt; Germany
- Funken, Horst; HHU-Düsseldorf; Jülich; Germany
- Furlan, Sara; National Research Council; Sesto Fiorentino (Firenze); Italy
- Galzitskaya, Oxana; Russian Academy of Science; Pushchino; Russia
- Gather, Malte; Universität zu Köln; Köln; Germany
- Gompper, Gerhard; Forschungszentrum Jülich; Jülich; Germany
- Griepernau, Beate; Saarland University; Saarbrücken; Germany
- Gront, Dominik; University of Virginia; Charlottesville; USA
- Gronwald, Wolfram; University of Regensburg; Regensburg; Germany
- Grudinin, Sergei; INRIA Grenoble; Saint-Ismier; France
- Hamacher, Kay; TU Darmstadt; Darmstadt; Germany
- Hansmann, Ulrich; Forschungszentrum Jülich; Jülich; Germany
- Hauser, Karin; University of Frankfurt; Frankfurt am Main; Germany
- Hayat, Sikander; University of Saarland; Saarbrücken; Germany
- Hayryan, Shura; Academis Sinica; Taipei; Taiwan (ROC)
- Helms, Volkhard; Saarland University; Saarbrücken; Germany
- Heuer, Andreas; University Münster; Münster; Germany

- Hoefling, Martin; LMU München; München; Germany
- Homeyer, Nadine; University of Erlangen; Erlangen; Germany
- Horn, Ruth; Fraunhofer IME; Aachen; Germany
- Hritz, Jozef; Vrije Universiteit; Amsterdam; The Netherlands
- Jamitzky, Ferdinand; High Performance Computing; Garching; Germany
- Johnson, Gary; Associate Director for Special Programs; Fort Collins, CO; USA
- Kar, Parimal; Michigan Technological University; Houghton, MI; USA
- Kartal, Önder; University of Potsdam; Potsdam-Golm; Germany
- Klenin, Konstantin; Universität Karlsruhe; Karlsruhe; Germany
- Klopffleisch, Karsten; Universität zu Köln; Köln; Germany
- Kmiecik, Sebastian; Selvita; Krakow; Poland
- Knecht, Volker; Max Planck Institute of Colloids and Interfaces; Potsdam; Germany
- Kobitski, Andrei; University of Ulm; Ulm; Germany
- König, Bernd; Forschungszentrum Jülich; Jülich; Germany
- Kokh, Daria; Bergische Universität Wuppertal; Wuppertal; Germany
- Kolinski, Andrzej; University of Warsaw; Warsaw; Poland
- König, Gerhard; University of Vienna; Vienna; Austria
- Krammer, Eva-Maria; Universität Bayreuth; Bayreuth; Germany
- Krause, Rolf; Universität Bonn; Bonn; Germany
- Krüger, Jens; National Yang-Ming University; Taipei; Taiwan
- Krüger, Peter; RWTH Aachen, Institut für Biochemie; Aachen; Germany
- Kubiak, Karina; University of Strathclyde; Glasgow; United Kingdom
- Kühn, Martin; Fraunhofer Institut für Techno- und Wirtschaftsmathematik; Kaiserslautern; Germany
- Kulczycka, Katarzyna; University of Warsaw; Warsaw; Poland
- Kumar, Subramanian Arun; University of Reading; Reading; United Kingdom
- La Penna, Giovanni; National Research Council; Sesto Fiorentino; Italy
- Law, Sean; Michigan State University; East Lansing, Michigan; USA
- Lenz, Jörn; University of Hamburg; Hamburg; Germany

- Lesyng, Bogdan; University of Warsaw, Faculty of Physics; Warsaw; Poland
- Li, Mai Suan; Polish Academy of Sciences; Warsaw; Poland
- Liepina, Inta; Latvian Institute of Organic Synthesis; Riga; Latvia
- Lippert, Thomas; Forschungszentrum Jülich; Jülich; Germany
- Liu, Binbin; University of Leeds; W. Yorkshire; United Kingdom
- Mahmood, Nasir; University of Hamburg; Hamburg; Germany
- Malnasi-Csizmadia, Andras; Eötvös University; Budapest; Hungary
- Margraf, Thomas; University of Hamburg; Hamburg; Germany
- Martínez, Leandro; Unité de Bioinformatique Structurale; Paris; France
- Meesters, Christian; University of Mainz; Mainz; Germany
- Meinke, Jan; Forschungszentrum Jülich; Jülich; Germany
- Meliciani, Irene; Forschungszentrum Karlsruhe; Eggenstein-Leopoldshafen; Germany
- Mohanty, Sandipan; Forschungszentrum Jülich; Jülich; Germany
- Nadler, Walter; Forschungszentrum Jülich; Jülich; Germany
- Nagel-Steger, Luitgard; Heinrich-Heine-Uni. Düsseldorf; Düsseldorf; Germany
- Namasivayam, Vigneshwaran; University of Leipzig; Leipzig; Germany
- Navarra-Madsen, Junalyn; TWU; Denton; USA
- Netz, Roland; Technical University Munich; München; Germany
- Neumann, Jan; LMU München; München; Germany
- Nöh, Katharina; Research Centre Jülich; Jülich; Germany
- Orland, Henri; Commissariat à l'Energie Atomique; Gif-sur-Yvette; France
- Orlowski, Slawomir; Nicolaus Copernicus University; Torun; Poland
- Pal, Sonia; Universität Köln; Köln; Germany
- Pálfi, Villõ; Eötvös University; Budapest; Hungary
- Peplowski, Lukasz; Nicolaus Copernicus University; Torun; Poland
- Pospiech, Christoph; IBM; Dresden; Germany
- Raaf, Jennifer; Universität zu Köln; Cologne; Germany
- Radestock, Sebastian; Christian-Albrechts-Universität; Kiel; Germany

- Reich, Lothar; Max-Planck-Gesellschaft; Potsdam; Germany
- Rice, John; IBM T.J. Watson Research Center; Yorktown Heights, NY; USA
- Rutkowska, Aleksandra; University of Warsaw; Warsaw; Poland
- Samsonov, Sergey; TU Dresden, BIOTEC; Dresden; Germany
- Sanchez, Horacio; Forschungszentrum Karlsruhe; Eggenstein-Leopoldshafen; Germany
- Schenk, Gundolf; Universität Hamburg; Hamburg; Germany
- Schiller, Annika; Forschungszentrum Jülich; Jülich; Germany
- Schnabel, Stefan; University of Leipzig; Leipzig; Germany
- Schomburg, Dietmar; Technische Universität Braunschweig; Germany
- Schröder, Walter; Forschungszentrum Jülich; Jülich; Germany
- Schug, Alexander; University of California San Diego; San Diego; USA
- Singh, Gurpreet; University of Dortmund; Dortmund; Germany
- Siu, Shirley Weng In; Saarland University; Saarbrücken; Germany
- Splett, Nicole; Universität zu Köln; Köln; Germany
- Strunk, Timo; Forschungszentrum Karlsruhe; Eggenstein-Leopoldshafen; Germany
- Stumpe, Martin; Max-Planck-Society; Göttingen; Germany
- Sutmann, Godehard; Research Centre Jülich; Jülich; Germany
- Suwattanasophon, Chonticha; University of Vienna; Vienna; Austria
- Szypowski, Marcin; University of Warsaw; Warsaw; Poland
- Tamamis, Phanourios; University of Cyprus; Nicosia; Cyprus
- Teyra, Joan; TU-Dresden, BIOTEC; Dresden; Germany
- Thielmann, Yvonne; Forschungszentrum Jülich; Jülich; Germany
- Thukral, Lipi; IWR; Heidelberg; Germany
- Tölzer, Christine; Universität Köln; Köln; Germany
- van Gunsteren, Wilfred F.; ETH Zürich; Zürich; Switzerland
- Vancraenenbroeck, Renée; KU Leuven; Heverlee; België
- Vogel, Thomas; University Leipzig; Leipzig; Germany
- Voltz, Karine; German Cancer Research Center; Heidelberg; Germany

- Wagner, Stephan; Universität zu Köln; Köln; Germany
- Weidemüller, Christian; J. W. Goethe-Universität; Frankfurt/Main; Germany
- Willbold, Dieter; Forschungszentrum Jülich; Jülich; Germany
- Wojciechowski, Michal; Polish Academy of Sciences; Warsaw; Poland
- Wolf, Maarten; Delft University of Technology; Delft; The Netherlands
- Yang, Wei; Florida State University; Tallahassee, FL; USA
- Zahran, Mai; University of Heidelberg; Heidelberg; Germany
- Zimmermann, Olav; Forschungszentrum Jülich; Jülich; Germany


Geb. 04.7
E1
04.7

Already published:

Modern Methods and Algorithms of Quantum Chemistry - Proceedings
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 1
ISBN 3-00-005618-1, February 2000, 562 pages
out of print

Modern Methods and Algorithms of Quantum Chemistry - Poster Presentations
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 2
ISBN 3-00-005746-3, February 2000, 77 pages
out of print

Modern Methods and Algorithms of Quantum Chemistry - Proceedings, Second Edition
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 3
ISBN 3-00-005834-6, December 2000, 638 pages
out of print

Nichtlineare Analyse raum-zeitlicher Aspekte der hirnelektrischen Aktivität von Epilepsiepatienten
Jochen Arnold
NIC Series Volume 4
ISBN 3-00-006221-1, September 2000, 120 pages

Elektron-Elektron-Wechselwirkung in Halbleitern: Von hochkorrelierten kohärenten Anfangszuständen zu inkohärentem Transport
Reinhold Lövenich
NIC Series Volume 5
ISBN 3-00-006329-3, August 2000, 146 pages

Erkennung von Nichtlinearitäten und wechselseitigen Abhängigkeiten in Zeitreihen
Andreas Schmitz
NIC Series Volume 6
ISBN 3-00-007871-1, May 2001, 142 pages

Multiparadigm Programming with Object-Oriented Languages - Proceedings
Kei Davis, Yannis Smaragdakis, Jörg Striegnitz (Editors)
Workshop MPOOL, 18 May 2001, Budapest
NIC Series Volume 7
ISBN 3-00-007968-8, June 2001, 160 pages

Europhysics Conference on Computational Physics - Book of Abstracts
Friedel Hossfeld, Kurt Binder (Editors)
Conference, 5 - 8 September 2001, Aachen
NIC Series Volume 8
ISBN 3-00-008236-0, September 2001, 500 pages

NIC Symposium 2001 - Proceedings
Horst Rollnik, Dietrich Wolf (Editors)
Symposium, 5 - 6 December 2001, Forschungszentrum Jülich
NIC Series Volume 9
ISBN 3-00-009055-X, May 2002, 514 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Lecture Notes
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 10
ISBN 3-00-009057-6, February 2002, 548 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms- Poster Presentations
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 11
ISBN 3-00-009058-4, February 2002, 194 pages

Strongly Disordered Quantum Spin Systems in Low Dimensions: Numerical Study of Spin Chains, Spin Ladders and Two-Dimensional Systems
Yu-cheng Lin
NIC Series Volume 12
ISBN 3-00-009056-8, May 2002, 146 pages

Multiparadigm Programming with Object-Oriented Languages - Proceedings
Jörg Striegnitz, Kei Davis, Yannis Smaragdakis (Editors)
Workshop MPOOL 2002, 11 June 2002, Malaga
NIC Series Volume 13
ISBN 3-00-009099-1, June 2002, 132 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Audio-Visual Lecture Notes
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 14
ISBN 3-00-010000-8, November 2002, DVD

Numerical Methods for Limit and Shakedown Analysis
Manfred Staat, Michael Heitzer (Eds.)
NIC Series Volume 15
ISBN 3-00-010001-6, February 2003, 306 pages

Design and Evaluation of a Bandwidth Broker that Provides Network Quality of Service for Grid Applications
Volker Sander
NIC Series Volume 16
ISBN 3-00-010002-4, February 2003, 208 pages

Automatic Performance Analysis on Parallel Computers with SMP Nodes
Felix Wolf
NIC Series Volume 17
ISBN 3-00-010003-2, February 2003, 168 pages

Haptisches Rendern zum Einpassen von hochaufgelösten Molekülstrukturdaten in niedrigaufgelöste Elektronenmikroskopie-Dichteverteilungen
Stefan Birmanns
NIC Series Volume 18
ISBN 3-00-010004-0, September 2003, 178 pages

Auswirkungen der Virtualisierung auf den IT-Betrieb
Wolfgang Gürich (Editor)
GI Conference, 4 - 5 November 2003, Forschungszentrum Jülich
NIC Series Volume 19
ISBN 3-00-009100-9, October 2003, 126 pages

NIC Symposium 2004
Dietrich Wolf, Gernot Münster, Manfred Kremer (Editors)
Symposium, 17 - 18 February 2004, Forschungszentrum Jülich
NIC Series Volume 20
ISBN 3-00-012372-5, February 2004, 482 pages

Measuring Synchronization in Model Systems and Electroencephalographic Time Series from Epilepsy Patients
Thomas Kreutz
NIC Series Volume 21
ISBN 3-00-012373-3, February 2004, 138 pages

Computational Soft Matter: From Synthetic Polymers to Proteins - Poster Abstracts
Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Editors)
Winter School, 29 February - 6 March 2004, Gustav-Stresemann-Institut Bonn
NIC Series Volume 22
ISBN 3-00-012374-1, February 2004, 120 pages

Computational Soft Matter: From Synthetic Polymers to Proteins - Lecture Notes
Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Editors)
Winter School, 29 February - 6 March 2004, Gustav-Stresemann-Institut Bonn
NIC Series Volume 23
ISBN 3-00-012641-4, February 2004, 440 pages

Synchronization and Interdependence Measures and their Applications to the Electroencephalogram of Epilepsy Patients and Clustering of Data
Alexander Kraskov
NIC Series Volume 24
ISBN 3-00-013619-3, May 2004, 106 pages

High Performance Computing in Chemistry
Johannes Grotendorst (Editor)
Report of the Joint Research Project:
High Performance Computing in Chemistry - HPC-Chem
NIC Series Volume 25
ISBN 3-00-013618-5, December 2004, 160 pages

Zerlegung von Signalen in unabhängige Komponenten: Ein informationstheoretischer Zugang
Harald Stögbauer
NIC Series Volume 26
ISBN 3-00-013620-7, April 2005, 110 pages

Multiparadigm Programming 2003
Joint Proceedings of the
3rd International Workshop on Multiparadigm Programming with Object-Oriented Languages (MPOOL'03)
and the
1st International Workshop on Declarative Programming in the Context of Object-Oriented Languages (PD-COOL'03)
Jörg Striegnitz, Kei Davis (Editors)
NIC Series Volume 27
ISBN 3-00-016005-1, July 2005, 300 pages

Integration von Programmiersprachen durch strukturelle Typanalyse und partielle Auswertung
Jörg Striegnitz
NIC Series Volume 28
ISBN 3-00-016006-X, May 2005, 306 pages

OpenMolGRID - Open Computing Grid for Molecular Science and Engineering
Final Report
Mathilde Romberg (Editor)
NIC Series Volume 29
ISBN 3-00-016007-8, July 2005, 86 pages

GALA Grünenthal Applied Life Science Analysis
Achim Kless and Johannes Grotendorst (Editors)
NIC Series Volume 30
ISBN 3-00-017349-8, November 2006, 204 pages

Computational Nanoscience: Do It Yourself!
Lecture Notes
Johannes Grotendorst, Stefan Blügel, Dominik Marx (Editors)
Winter School, 14. - 22 February 2006, Forschungszentrum Jülich
NIC Series Volume 31
ISBN 3-00-017350-1, February 2006, 528 pages

NIC Symposium 2006 - Proceedings
G. Münster, D. Wolf, M. Kremer (Editors)
Symposium, 1 - 2 March 2006, Forschungszentrum Jülich
NIC Series Volume 32
ISBN 3-00-017351-X, February 2006, 384 pages

Parallel Computing: Current & Future Issues of High-End Computing
Proceedings of the International Conference ParCo 2005
G.R. Joubert, W.E. Nagel, F.J. Peters,
O. Plata, P. Tirado, E. Zapata (Editors)
NIC Series Volume 33
ISBN 3-00-017352-8, October 2006, 930 pages

From Computational Biophysics to Systems Biology 2006 Proceedings
U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann (Editors)
NIC Series Volume 34
ISBN-10 3-9810843-0-6, ISBN-13 978-3-9810843-0-6,
September 2006, 224 pages

Dreistufig parallele Software zur Parameteroptimierung von Support-Vektor-Maschinen mit kostensensitiven Gütemaßen
Tatjana Eitrich
NIC Series Volume 35
ISBN 978-3-9810843-1-3, March 2007, 262 pages

From Computational Biophysics to Systems Biology (CBSB07) Proceedings
U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann (Editors)
NIC Series Volume 36
ISBN 978-3-9810843-2-0, August 2007, 330 pages

Parallel Computing: Architectures, Algorithms and Applications - Book of Abstracts
Book of Abstracts, ParCo 2007 Conference, 4. - 7. September 2007
G.R. Joubert, C. Bischof, F. Peters, T. Lippert, M. Bücker, P. Gibbon, B. Mohr (Eds.) NIC Series Volume 37
ISBN 978-3-9810843-3-7, August 2007, 216 pages

Parallel Computing: Architectures, Algorithms and Applications - Proceedings
Proceedings, ParCo 2007 Conference, 4. - 7. September 2007
C. Bischof, M. Bücker, P. Gibbon, G.R. Joubert, T. Lippert, B. Mohr, F. Peters (Eds.) NIC Series Volume 38
ISBN 978-3-9810843-4-4, December 2007, 830 pages

NIC Symposium 2008 - Proceedings
G. Münster, D. Wolf, M. Kremer (Editors)
Symposium, 20 - 21 February 2008, Forschungszentrum Jülich
NIC Series Volume 39
ISBN 978-3-9810843-5-1, February 2008, 380 pages

All volumes are available online at

http:// www.fz-juelich.de/nic-series/.