From Computational Biophysics to Systems Biology (CBSB11) – Celebrating Harold Scheraga's 90th Birthday

Proceedings, 20 – 22 July 2011 | Jülich, Germany

Paolo Carloni, Ulrich H.E. Hansmann, Thomas Lippert, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, Olav Zimmermann (Editors)

JÜLICH
FORSCHUNGSZENTRUM

From Computational Biophysics to Systems Biology (CBSB11) – Celebrating Harold Scheraga's 90th Birthday

Proceedings, 20 – 22 July 2011 | Jülich, Germany

edited by Paolo Carloni, Ulrich H.E. Hansmann, Thomas Lippert, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, Olav Zimmermann

Preface

About 110 researchers and students from all over the world met from July 20th to July 22nd at the Forschungszentrum Jülich to participate in the international workshop "From Computational Biophysics to Systems Biology (CBSB11)". The workshop was dedicated to Harold Scheraga who celebrates his 90th birthday in October 2011. Dr. Scheraga pioneered the use of computers in chemistry and biology. His work inspired many of the research areas that were the topic of this meeting, ranging from biophysics to systems biology. In the spirit of Harold Scheraga's work, we have aimed at bringing together researchers from physics, chemistry, biology, and computer science to acquaint each other with current trends in computational biophysics and systems biology, to explore avenues of cooperation, and to establish together a detailed understanding of cells at a molecular level.

CBSB11 was the fifth in a series of successful workshops held since 2006 in both Germany and the USA, and was organized jointly by the Forschungszentrum Jülich, the German Research School for Simulation Sciences, and Michigan Technological University. As in the previous year, the participants explored a wide range of topics ranging from single macromolecules to the working of entire cells in scientific presentations (including five "CBSB11 Outstanding Young Researcher Award" talks by selected students and postdocs), in panel sessions, and in numerous informal discussions. This proceeding volume collects selected presentations from the three days long workshop that we hope will serve as starting point for further discussions.

We are indebted to Bettina Scheid, Elke Bielitza and Helga Frank for their most valuable help with local arrangements. Martina Kamps helped not only with organizing the workshop, but also maintained the webpage and helped with editing these proceedings. We also wish to thank IBM for support.

Jülich, September 2011

Paolo Carloni
Ulrich Hansmann
Thomas Lippert
Jan Meinke
Sandipan Mohanty
Walter Nadler
Olav Zimmermann

Contents

Laudation

Key Note Speaker

Talks and Posters

A Tribute to the First 90 Years of Harold A. Scheraga

Jeffrey Skolnick

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology
250 14th Street, N.W., Atlanta, GA 30318, USA
E-mail: skolnick@gatech.edu

Harold A Scheraga was born on October 18, 1921 to a working class family in Brooklyn, New York. He then spent the early part of his childhood in Monticello, NY, before returning to Brooklyn in 1930. Typical of many in those days, during the Depression, he almost dropped out of public school to help support his family. But, fortunately for the scientific world, due to the influence of his father, he remained in school. Harold has always had a keen interest in the humanities. For example, in Junior High School, he extensively studied Latin, where he was award the Latin Medal on graduation. He then attended the prestigious Boys High School, where he also developed a passion for math and was a member of the math team; yet, during this time, his love of Latin and the classics increased. Indeed, on entering the City College of New York, he was torn between the arts and math. Ultimately, he decided that chemistry was to be his field of study and took courses in both math and physics to prepare himself for a career in physical chemistry. Tellingly, he avoided Biochemistry, because "I didn't think it was a rigorous scientific discipline. I've since done a 180° phase shift on that one."

In 1941, Harold Scheraga went on to graduate school in chemistry at Duke University, where he studied the Kerr effect in small molecules under the direction of Paul Gross. On completion of his PhD, he was awarded a prestigious ACS Postdoctoral Fellowship in 1946-47 that enabled him to work with John Edsall at Harvard Medical School. There he worked on fibronectin and tried to determine its size and shape using dielectric dispersion and flow birefringence. This marked the beginning of his lifelong interest in protein biophysics.

In 1947, he was appointed as a Chemistry Instructor at Cornell University by Peter Debye and quickly worked his way up the ranks, becoming a Full Professor in 1958 and the chair of the Cornell Chemistry Department from 1960-1967. It was under his chairmanship that the Cornell Chemistry Department underwent a period of rapid growth. In 1965, he was appointed the Todd Professor of Chemistry, a position he held until 1992, when he became the Todd Professor Emeritus, and to this day, he continues to direct a vigorous research program.

In a rarely equaled, scientific career spanning over 65 years, Dr. Scheraga has been and remains one of the world's leading protein biophysical chemists. In a unique combination of experiment and theory, he has made an exceptionally large number of original and pioneering contributions to the development and application of physical and chemical methods to polypeptides and proteins. Dr. Scheraga was the very first person that conceived of applying the fundamental principles of physical chemistry to the problem of protein structure prediction. As the author of over 1270 scientific articles, he has introduced almost all of the key ideas that have led to the understanding of protein stability, the

mechanism of protein folding and the prediction of protein structure. Indeed, four of his papers are citation classics.

Among his major scientific contributions are the following: He developed a method to interpret the hydrodynamic properties of proteins. He developed a theory of the hydrophobic effect. He engaged in a pioneering analysis of phase transitions in one dimension, and in related work, studied the helix-coil transition in homo- and copolymers of amino acids. He then applied the host-guest technique to determine the intrinsic helix-forming tendencies of the naturally occurring amino acids with modifications to include the role of specific side chain-side chain and side chain-backbone interactions. He engaged in an ingenious series of chemical and physical experiments to obtain structurally relevant distance constraints in ribonuclease. These were subsequently verified by the crystal structure of ribonuclease. Not merely interested in ribonuclease's tertiary structure, he wondered how it folds to its native state; thus, he determined the folding pathways of ribonuclease A. He also worked on β-hairpin-forming peptides as models of early stages of protein folding. Following up on his postdoctoral work on fibronectin, he elucidated the mechanism of the thrombin-induced conversion of fibrinogen to fibrin including the identification of the molecular basis of a bleeding disorder. He performed the first calculations of the structures of enyzme-substrate complexes. Moreover, he is one of the founders of the field of conformational energy calculations. In that regard, he developed the ECEPP force field for proteins and is the author of numerous highly creative algorithms for the solution of the multiple-minima problem for oligopeptides, fibrous proteins, and globular proteins. Indeed, his diffusion-equation and distance-scaling methods of global optimization spawned an entire field of energy surface deformations. He subsequently applied these approaches to compute crystal structures solely from the potential function with no input of space-group information. This was the first time such first principle calculations were successful. He was also the first to introduce statistical potentials for use in protein structure prediction. Focusing on fundamental physical principles, he elucidated the fundamental statistical mechanical origins of cooperativity in protein folding and he assessed the accuracy of protein structures by applying quantum mechanics to the calculation of $^{13}C^{\alpha}$ chemical shifts. Most recently, he has pioneered the development of a reduced protein model, UNRES, that has enabled the first principle simulation of the folding pathway of a number of important proteins that represents significant progress towards the solution of the protein folding problem.

In recognition of his outstanding science, he has received numerous honors throughout his career. He was a Guggenheim Fellow and Fulbright Research Scholar at the Carlsberg Lab, Copenhagen, Denmark, 1956-57 and at the Weizmann Institute in 1963. He was an NIH Health Special Fellow at the Weizmann Institute in 1970, and an NIH Fogarty Scholar in 1984, 1986, 1988, 1989, 1990, and 1991. In 1957, he received the ACS Eli Lilly Award in Biochemistry. He was awarded a Sc.D. (Hon.) from Duke University in 1961 and the University of Rochester in 1988. He was elected a Fellow of the American Association for the Advancement of Science in 1966, a Member of the National Academy of Sciences, U.S. in 1966 and a Member of the American Academy of Arts and Sciences in 1967. In 1978, he received the ACS Kendall Award in Colloid or Surface Chemistry, and in 1983, he was awarded the Linderstrøm-Lang Medal of Carlsberg Laboratory. In 1990, he received the ACS Mobil Award in Polymer Chemistry and the ACS Repligen Award for Chemistry of Biological Processes. He was also awarded a Doctor Scientiarum Honoris Causa from the

Technion, Israel Institute of Technology in 1993. In 1995, he received the Stein and Moore Award of the Protein Society, and in 1997, he was honored by the ACS IBM Award for Computers in Chemical and Pharmaceutical Research. In 2009, he was awarded the ACS Murray Goodman Award and was elected a Fellow of the American Chemical Society.

Harold Scheraga has also had a highly significant impact on the scientific community in his role as mentor of over 300 graduate students and postdoctoral fellows. Many of these students have gone on to outstanding careers in academia and industry. His dedication to mentoring the next generations of scientists is also reflected in his choice of speakers at his numerous awards symposia honoring his accomplishments. He always selects young people who are at the cutting edge of their science.

At the young age of 90, Harold Scheraga maintains a vigorous passion for cutting edge science. Indeed, his NIH grant, which is in its 55th year, was renewed with the highest priority score. Harold is well known for often asking the most relevant, incisive question at a seminar and for pointing out that he did the first work on the problem, perhaps some 40 years earlier. What is even more remarkable is that he can, off the top of his head, give you the exact citation of that work! This is a non-trivial task, considering the number of papers he has published and continues to publish.

The scientific community is very fortunate to have Harold as one of its senior members. He sets a high standard not only in his world-class science but also in his world-class ethical behavior. As shown by the diverse collection of papers in this volume, his science has played a seminal role that has spread across many fields of biology, and we all look forward to his many future contributions in the years to come.

Adventures in Protein Biophysics

Harold A. Scheraga

Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301, USA
E-mail: has5@cornell.edu

Highlights of 65 years in Protein Biophysics are presented. The science evolved along experimental and theoretical pathways from early hydrodynamic treatments of protein solutions to all-atom, and then coarse-grained, molecular mechanics simulations of protein structure and thermodynamics of folding pathways. Finally, applications are now being made to treat several biological systems.

1   Introduction

This is a summary of some of my 65 years in Protein Biophysics, during which I have had the good fortune to establish friendships with wonderful people: my students and many others throughout the world including several people participating in this symposium. The following table is a basic description of my dual experimental and theoretical approach to determine protein structure and reactivity.

<div align="center">

Dual Biophysics Approach to Determine
Protein Structure and Reactivity

</div>

Experimental		Theoretical
Hydrodynamic properties to determine size and shape of ellipsoidal models	$\longrightarrow$	Hydrodynamic theory to determine rotational diffusion coefficients of ellipsoidal molecules
		Generalized treatment of hydrodynamic properties of proteins
Mechanism of thrombin-induced conversion of fibrinogen to fibrin (hydrogen-bonding in protein-protein association)		Pauling α and β structures based on backbone hydrogen-bonding
		$\downarrow$
Use of random copolymers to determine helix-coil propensity (Host-Guest technique), and failure of this technique to account for phase transitions in proteins	$\longleftarrow$	Helix-coil transition theory for polyamino acids and nucleic acids
		$\downarrow$

	Theory of effect of hydrogen-bonds on pK's of polar side chains and on limited proteolysis
	Similar theory for nonpolar side chains (hydrophobic interaction)
Structure of proteins, making use of side-chain distance constraints (Tyr...Asp) ⟶	Formulation of molecular mechanics treatment to compute protein structure
↓	↓
Mechanism of oxidative folding of RNase A and onconase (Determination of folding pathways)	Evolved into ECEPP and UNRES structure determination (CASP)
	↓
	Later treatment of thermodynamics and dynamics
	↓
New venture in exploiting use of $^{13}C^{\alpha}$ chemical shifts	Protein – Protein Interactions (Biological properties)
	↓
	Mean-square fluctuations in native proteins, and energy landscapes (PCA)

2 Hydrodynamic Properties

In 1946, other than that a protein contained peptide-bond-linked amino acids, essentially the only structural property of a protein that could be determined was its size and shape using hydrodynamic methods. My first experimental experience with proteins as a post-doc in the Department of Physical Chemistry at Harvard Medical School involved flow birefringence to determine the dimensions of an ellipsoidal model for a particular protein then-known as cold-insoluble globulin[1]. This was followed by theoretical work to obtain this information from rotational diffusion coefficients of ellipsoidal molecules from flow birefringence[2], and subsequently, from non-Newtonian viscosity[3] experiments. Later, be-

ginning at Cornell in 1947, a general treatment of a variety of hydrodynamic properties was formulated[4] with Leo Mandelkern.

3 Thrombin-induced Conversion of Fibrinogen to Fibrin

Simultaneously with flow-birefringence studies of fibrinogen[5] and detergent micelles[6], a long-time experimental study was initiated with my first graduate student, Michael Laskowski, Jr., to determine the mechanism of the interaction of thrombin with fibrinogen to produce the fibrin clot[7,8]. The following reversible three-step mechanism was elucidated:

$$F \overset{T}{\Leftrightarrow} P + f \tag{1}$$

$$nf \Leftrightarrow f_n \tag{2}$$

$$mf_n \Leftrightarrow \text{fibrin clot} \tag{3}$$

in which F is the asymmetrical rod-like molecule, fibrinogen, T is the proteolytic enzyme thrombin, P is small peptide material liberated by hydrolysis of a specific Arg16-Gly17 peptide bond by thrombin to expose buried polymerization sites on F, producing fibrin monomer, f, in step 1.

In step 2, this exposed polymerization site on F interacts with a pre-existing site on F to produce a distribution of oligomeric intermediate over-lapping rod-like polymers[7,9], f_n. This polymerization is facilitated by formation of intermolecular hydrogen bonds[7,10].

Finally, in step 3, these rod-like intermediate polymers form a cross-linked gel whose characteristics were described by Ferry and Morrison[11]. Subsequent transferred nuclear Overhauser effect (NOE) NMR experiments identified the structure of the initial complex[12] between T and F, as shown by Fig. 7 in Ref. 12. In a genetic mutation in some patients, the nearby Gly12 is replaced by valine, altering the conformation of F near the active-site Arg16-Gly17 bond, preventing the hydrolysis of this peptide bond and thereby avoiding the formation of f or the fibrin clot, leading to a bleeding disorder. With a transferred NOE NMR experiment with this mutant fibrinogen, the molecular mechanism of this bleeding disorder was elucidated[13].

4 Helix-coil Transitions

In the early 1950's, Pauling proposed the α-helix and parallel and anti-parallel β sheets as structural elements of proteins. This started extensive investigations in many labs including ours with a new graduate student, Douglas Poland, of the helix-coil transition. Our goal was to determine the helix-forming propensity of the 20 naturally-occurring amino-acid residues. While it would have been desirable to obtain this information from equilibrium studies of the helix-coil transition in homopolymers of each of the 20 amino acids in water, most of them are insoluble in this solvent. Therefore, synthetic host-guest random copolymers, with a water-soluble host doped, respectively, with each of the 20 amino acids singly as a guest, was used[14], and an equilibrium theory was developed[15,16] for helix-coil transitions in random copolymers. The helix propensity was expressed in terms of the Zimm-Bragg equilibrium constant, s, for breaking a hydrogen bond at the end of the α-helix, and s was determined over the temperature range of $0°$ to $60°C$ for all 20 amino-acid

residues[17,18]. Such s-values are useful for identifying initial formation of helical structures in protein-folding studies[19,20]. At the same time, a theory was developed to treat the kinetics of the helix-coil transition in polyamino acids[21].

A large focus was placed on helix-coil transitions because it was widely believed that this process could ultimately provide an understanding of protein folding. However, it was later realized that such a concept was incorrect because the nearest-neighbor Ising model, used to treat the helix-coil transition, did not include the proper balance between short- and long-range interactions that governs cooperative folding of proteins[22].

5   Hydrogen Bonding between Side Chains

Pauling's proposal of α and β structures focused on NH$\cdots$OC hydrogen bonds in the backbone of the polypeptide chain. Stimulated by the identification of hydrogen bonds between side chains in the polymerization of fibrin monomer[7,10], a theory was developed with Michael Laskowski, Jr. for the effect of such hydrogen bonding of polar side chains on pK's of ionizable groups[23] and on limited proteolysis[24]. Correspondingly, theory was proposed with George Némethy for the mechanism of hydrophobic interactions involving nonpolar side chains[25], and for the interaction of nonpolar residues with the nonpolar portions of polar amino acids to strengthen hydrogen bonds in which the polar side chains are involved[26]; the latter enhanced our understanding of the contributions of such hydrogen bonds to protein stability.

6   Hydrogen-bond Distance Constraints to Determine Protein Structure

With the ability to identify side chain-side chain hydrogen-bonds between polar residues by their effect on pK's[23], attempts were made to identify such hydrogen-bond distance constraints in a protein to facilitate determination of its three-dimensional structure. Using bovine pancreatic ribonuclease A (RNase A) as an example, three such hydrogen bonds were identified in RNase A[27], viz., Tyr 25$\cdots$Asp 14, Tyr 92...Asp 38, and Tyr 97...Asp 83, before the crystal structure had been determined. This pairing of 3 out of 6 tyrosine residues with 3 out of 11 carboxyl groups represents one out of more than 19,000 possible combinations. When the crystal structure of RNase A was subsequently determined, these three pairs were found. These three distance constraints, plus the known location of four disulfide bonds in RNase A encouraged us to start to develop a theoretical approach with George Némethy, using distance constraints[28], ultimately with inclusion of empirical potentials[29], to compute the three-dimensional structure of a protein.

7   Experimental Studies of Protein-folding Pathways

The experimental work on identifying hydrogen bonds in RNase A[27] was accompanied by an experimental approach to identify the pathways from the unfolded form of this protein (with its disulfide bonds reduced) to the oxidatively-folded protein to form its four native disulfide bonds. These oxidative-folding experiments were carried out on RNase A, first,

with oxidized-and-reduced glutathione[30] but, later, to simplify the number of folding intermediates, with oxidized and reduced dithiothreitol[31,32]. Two pathways, in which the rate-determining step involved SH/S-S interchange to form the three-disulfide-bonded species des-[65-72] and des-[40-95], were identified[31]. Two other pathways, in which these two disulfide-bonded species are formed by oxidation from the two-disulfide ensemble, rather than by SH/S-S interchange, were also identified[33,34]. In the initial folding stage, the one-disulfide ensemble is dominated by formation of the 65-72 disulfide bond[35].

A ribonuclease homolog, onconase (ONC), also has four disulfide bonds and a three-dimensional structure similar to that of RNase A. Three of the disulfide bonds of ONC and RNase A are in homologous positions, but the fourth one in RNase A, between Cys 65 and Cys 72, which is the first disulfide bond to form in the folding pathways, is in a different, non-homologous, position in ONC. This motivated an analogous study of the oxidative folding of ONC[36] with dithiothreitol, which indeed led to different folding pathways than those observed in RNase A. Thus, these two homologous proteins fold by very different pathways.

8 Formulation of Molecular Mechanics to Compute Protein Structure

With the distance constraints mentioned in section 6, and a hard-sphere potential, and constraints of the Ramachandran diagram[37], methodology was formulated[28] to compute and assess the accessibility of various folded states of RNase A. This methodology evolved into use of a detailed empirical potential energy function, ECEPP (Empirical Conformational Energy Program for Peptides)[29].

Several small model peptides were investigated to demonstrate the validity of the molecular mechanics procedure, with various global optimization procedures[38], to surmount the many local minima in the conformational space and produce structures that agreed with experimental data. These include the pentapeptide methionine enkephalin, the cyclic decapeptide gramicidin S, a polytripeptide model of collagen, and the 46-residue protein A.

Protein A was the largest protein that could be treated with the available computing power[39]. To be able to compute structures of larger proteins than protein A, a united-residue (UNRES) coarse-grained model[40] was developed by Adam Liwo by eliminating unnecessary degrees of freedom to limit the conformational space that has to be examined.

UNRES evolved from an initial dependence on the protein data bank for parameterization to a more physics-based one parameterized on *ab initio* quantum mechanical calculations and computations of potentials of mean force[40]. Its initial success with single-chain proteins in CASP involved the computation of the structure of the protein HDEA[41]. With such success in computing structure, attention was turned to the use of UNRES in a molecular dynamics treatment of folding pathways of single-chain proteins[42–44] to be able to extend the computational folding time-scale beyond the one involved with all-atom potentials. An extension of UNRES to multiple-chain proteins was carried out subsequently[45].

9   Thermodynamics and Dynamics

To correct for the focus on energy instead of free energy, the UNRES force field was modified to include its temperature dependence[46], and thereby treat entropic effects to compute thermodynamic properties. In order to extend UNRES to proteins containing as many as 500 residues, massive parallelization of the force field was carried out[47]. At the same time, attention began to be focused on dynamical fluctuations, initially of native proteins[48,49].

10   Biological Processes

With the increasing maturity of UNRES, this force field was applied to treat large biological protein-protein complexes. The initial focus was on the aggregation of Aβ[50] and PICK1[51].

11   Use of $^{13}C^{\alpha}$ Chemical Shifts

With Jorge Vila, we have recently initiated the computation and application of NMR $^{13}C^{\alpha}$ chemical shifts to determine and validate protein structures. Quantum chemical calculations were carried out for any residue Xxx in a protein with the use of an ECEPP-rigid-geometry model peptide Ac-Gly-Xxx-Gly-NMe. The factors that affect the DFT-based computation of $^{13}C^{\alpha}$ chemical shifts were established[52–54], and it was shown that complete protein structures solved by X-ray crystallography and NMR spectroscopy could be validated accurately[54–57]. In addition, the fractions of the tautomeric forms of the imidazole ring of histidine in proteins could be determined as a function of pH[58]. As a result, a purely physics-based server, *Che*Shift[59], was developed to enable the NMR community to predict of $^{13}C^{\alpha}$ chemical shifts with reasonable accuracy in seconds; thus, this provides a standard with which to evaluate the quality of any reported protein structure solved by either X-ray crystallography or NMR spectroscopy, provided that the experimentally-observed $^{13}C^{\alpha}$ chemical shifts are available.

12   Conclusions

Methodology has been developed to compute protein structure and folding pathways, together with the thermodynamic and dynamic properties during folding. Current focus is being placed on the improvement of UNRES and its application to biological protein-protein complexes, e.g., chaperones. More extensive details are provided in a recently-published chapter in Annual Review of Biophysics[60].

References

1. J. T. Edsall, G. A. Gilbert and H. A. Scheraga. The Nonclotting component of the human plasma fraction I1 ("Cold Insoluble Globulin"). J. Am. Chem. Soc., 77: 157-161 (1955).

2. H. A. Scheraga, J. T. Edsall and J. O. Gadd, Jr. Double refraction of flow: Numerical evaluation of extinction angle and birefringence as a function of velocity gradient. J. Chem. Phys., 19: 1101-1108 (1951).

3. H. A. Scheraga. NonNewtonian viscosity of solutions of ellipsoidal particles. J. Chem. Phys., 23: 1526-1532 (1955).

4. H. A. Scheraga and L. Mandelkern. On the interpretation of hydrodynamic data for dilute protein solutions. J. Phys. Chem., 62: 370 (1958).

5. C. S. Hocking, M. Laskowski, Jr. and H. A. Scheraga. Size and shape of bovine fibrinogen. J. Am. Chem. Soc., 74: 775-778 (1952).

6. H. A. Scheraga and J. K. Backus. Flow birefringence in solutions of n-hexadecyltrimethylammonium bromide. J. Am. Chem. Soc., 73: 5108-5112 (1951).

7. H. A. Scheraga and M. Laskowski, Jr. The fibrinogen-fibrin conversion. Adv. Prot. Chem., 12: 1-131 (1957).

8. H.A. Scheraga. The thrombin-fibrinogen interaction. Biophys. Chem., 112: 117-130 (2004).

9. T. H. Donnelly, M. Laskowski, Jr., N. Notley and H. A. Scheraga. Equilibria in the fibrinogen fibrin conversion. II. Reversibility of the polymerization steps. Arch. Biochem. and Biophys., 56: 369-387 (1955).

10. J. M. Sturtevant, M. Laskowski, Jr., T. H. Donnelly and H. A. Scheraga. Equilibria in the fibrinogenfibrin conversion. III. Heats of polymerization and clotting of fibrin monomer. J. Am. Chem. Soc., 77: 6168-6172 (1955).

11. J.D. Ferry and P.R. Morrison. Preparation and properties of serum and plasma proteins. VIII. The conversion of human fibrinogen to fibrin under various conditions. J. Am. Chem. Soc., 69: 388-400 (1947).

12. F. Ni, Y. C. Meinwald, M. Vasquez and H. A. Scheraga. Highresolution NMR studies of fibrinogenlike peptides in solution: Structure of a thrombin bound peptide corresponding to residues 716 of the Aα chain of human fibrinogen, Biochemistry, 28: 3094-3105 (1989).

13. F. Ni, Y. Konishi, L. D. Bullock, M. N. Rivetna and H. A. Scheraga. Highresolution NMR studies of fibrinogenlike peptides in solution: Structural basis for the bleeding disorder caused by a single mutation of Gly(12) to Val(12) in the Aα chain of human fibrinogen Rouen. Biochemistry, 28: 3106-3119 (1989).

14. P. H. Von Dreele, N. Lotan, V. S. Ananthanarayanan, R. H. Andreatta, D. Poland and H. A. Scheraga. Helixcoil stability constants for the naturally occurring amino acids in water. II. Characterization of the host polymers and application of the host-guest technique to random poly(hydroxypropylglutamine-co-hydroxybutyl-glutamine). Macromolecules, 4: 408-417 (1971).

15. D. Poland and H. A. Scheraga. The LifsonAllegra theories of the helixcoil transition for random copolymers: comparison with exact results and extension. Biopolymers, 7: 887-908 (1969).

16. P. H. Von Dreele, D. Poland and H. A. Scheraga. Helix-coil stability constants for the naturally occurring amino acids in water. I. Properties of copolymers and approximate theories. Macromolecules, 4: 396-407 (1971).

17. J. Wojcik, K. H. Altmann and H. A. Scheraga. Helix coil stability constants for the naturally occurring amino acids in water. XXIV. Half cystine parameters from random Poly(hydroxybutylglutamine-co-S-methylthio-L-cysteine). Biopolymers, 30: 121-134 (1990).

18. H.A. Scheraga, J.A. Vila and D.R. Ripoll. Helix-coil transitions re-visited. Biophys., Chem., 101-102:255-265 (2002).

19. P. N. Lewis, N. Gō, M. Gō, D. Kotelchuck and H. A. Scheraga. Helix probability profiles of denatured proteins and their correlation with native structures. Proc. Natl. Acad. Sci., U.S., 65: 810-815 (1970).

20. P. N. Lewis and H. A. Scheraga. Prediction of structural homology between bovine α-lactalbumin and hen egg white lysozyme. Arch. Biochem. Biophys., 144: 584-588 (1971).

21. D. Poland and H. A. Scheraga. Kinetics of the helixcoil transition in polyamino acids. J. Chem. Phys., 45: 2071-2090 (1966).

22. M.-H. Hao and H. A. Scheraga. Theory of two-state cooperative folding of proteins. Acct. Chem. Res., 31: 433-440 (1998).

23. M. Laskowski, Jr. and H. A. Scheraga. Thermodynamic considerations of protein reactions. I. Modified reactivity of polar groups. J. Am. Chem. Soc., 76: 6305-6319 (1954).

24. M. Laskowski, Jr. and H. A. Scheraga. Thermodynamic considerations of protein reactions. II. Modified reactivity of primary valence bonds. J. Am. Chem. Soc., 78: 5793-5798 (1956).

25. G. Némethy and H. A. Scheraga. The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. J. Phys. Chem., 66: 1773-1789 (1962). Erratum: J. Phys. Chem., 67: 2888 (1963).

26. G. Némethy, I. Z. Steinberg and H. A. Scheraga. The influence of water structure and of hydrophobic interactions on the strength of side-chain hydrogen bonds in proteins. Biopolymers, 1: 43-69 (1963).

27. H. A. Scheraga. Structural studies of pancreatic ribonuclease. Fed. Proc., 26: 1380-1387 (1967).

28. G. Nmethy and H. A. Scheraga. Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method, Biopolymers, 3: 155-184 (1965).

29. F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem., 79: 2361-2381 (1975).

30. Y. Konishi, T. Ooi and H. A. Scheraga. Regeneration of RNase A from the reduced protein: models of regeneration pathways, Proc. Natl. Acad. Sci., U.S., 79: 5734-5738 (1982).

31. D. M. Rothwarf, Y.-J. Li and H. A. Scheraga. Regeneration of bovine pancreatic ribonuclease A. Identification of two nativelike three-disulfide intermediates involved in separate pathways. Biochemistry, 37: 3760-3766 (1998).

32. D. M. Rothwarf, Y.-J. Li and H. A. Scheraga. Regeneration of bovine pancreatic ribonuclease A. Detailed kinetic analysis of two independent folding pathways. Biochemistry, 37: 3767-3776 (1998).

33. M. Iwaoka, D. Juminaga and H. A. Scheraga. Regeneration of three-disulfide mutants of bovine pancreatic ribonuclease A missing the 65-72 disulfide bond: Characterization of a minor folding pathway of ribonuclease A and kinetic roles of Cys65 and Cys72. Biochemistry, 37: 4490-4501 (1998).

34. X. Xu and H. A. Scheraga. Kinetic folding pathway of a three-disulfide mutant of bovine pancreatic ribonuclease A missing the [40-95] disulfide bond. Biochemistry, 37: 7561-7571 (1998).

35. X. Xu, D. M. Rothwarf and H. A. Scheraga. Nonrandom distribution of the one-disulfide intermediates in the regeneration of ribonuclease A. Biochemistry, 35: 6406-6417 (1996).

36. R.F. Gahl and H.A. Scheraga. Oxidative folding pathway of onconase, a ribonuclease homologue: Insight into oxidative folding mechanisms from a study of two homologues. Biochemistry, 48: 2740-2751 (2009).

37. G.N. Ramachandran, C. Ramakrishnan and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7: 95-99 (1963).

38. H. A. Scheraga. Recent developments in the theory of protein folding: searching for the global energy minimum, Biophys. Chem. 59: 329-339 (1996).

39. J.A. Vila, D.R. Ripoll and H.A. Scheraga. Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures. Proc. Natl. Acad. Sci., U.S.A., 100: 14812-14816 (2003).

40. A. Liwo, C. Czaplewski, S. Ołdziej, A.V. Rojas, R. Kazmierkiewicz, M. Makowski, R.K. Murarka and H.A. Scheraga. Simulation of protein structure and dynamics with the coarse-grained UNRES force field, in "Coarse-Graining of Condensed Phase and Biomolecular Systems", ed. G.A. Voth, CRC Press, pp. 107-122 (2008).

41. A. Liwo, J. Lee, D. R. Ripoll, J. Pillardy and H. A. Scheraga. Protein structure prediction by global optimization of a potential energy function, Proc. Natl. Acad. Sci., USA, 96: 5482-5485 (1999).

42. A. Liwo, M. Khalili and H.A. Scheraga. *Ab initio* simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. Proc. Natl. Acad. Sci., U.S.A., 102: 2362-2367 (2005).

43. M. Khalili, A. Liwo, F. Rakowski, P. Grochowski and H.A. Scheraga. Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode, J. Phys. Chem. B., 109: 13785-13797 (2005).

44. M. Khalili, A. Liwo, A. Jagielska and H.A. Scheraga. Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model α-helical systems. J. Phys. Chem. B, 109: 13798-13810 (2005).

45. A. V. Rojas, A. Liwo and H.A. Scheraga. Molecular dynamics with the united-residue (UNRES) force field. *Ab initio* folding simulations of multi-chain proteins. J. Phys. Chem. B., 111: 293-309 (2007).

46. A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik and H.A. Scheraga. Modification and optimization of the united-residue (UNRES) potential-energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. J. Phys. Chem. B., 111: 260-285 (2007).

47. A. Liwo, S. Ołdziej, C. Czaplewski, D.S. Kleinerman, P. Blood and H.A. Scheraga. Implementation of molecular dynamics and its extensions with the coarse-grained UNRES force field on massively parallel systems; towards millisecond-scale simulations of protein structure, dynamics, and thermodynamics. J. Chem. Theory and Comput., 6: 890-909 (2010).

48. P. Senet, G.G. Maisuradze, C. Foulie, P. Delarue, and H.A. Scheraga. How main-chains of proteins explore the free-energy landscape in native states. Proc. Natl. Acad. Sci., U.S.A., 105: 19708-19713 (2008).

49. Y. Cote, P. Senet, P. Delarue, G.G. Maisuradze and H.A. Scheraga. Nonexponential decay of internal rotational correlation functions of native proteins and self-similar structural fluctuations. Proc. Natl. Acad. Sci., U.S.A., 107: 19844-19849 (2010).

50. A. Rojas, A. Liwo, D. Browne and H.A. Scheraga. Mechanism of fiber assembly; treatment of $A\beta$-peptide aggregation with a coarse-grained united-residue force field. J. Mol. Biol., 404: 537-552 (2010).

51. Y. He, A. Liwo, H. Weinstein and H.A. Scheraga. PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics. J. Mol. Biol., 405: 298-314 (2011).

52. 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: Structure, Function, and Bioinformatics, 71: 641-654 (2008).

53. J.A. Vila, H.A. Baldoni and H.A. Scheraga. Performance of density functional models to reproduce observed $^{13}C^{\alpha}$ chemical shifts of proteins in solution. J. Comput. Chem., 30: 884-892 (2009).

54. J.A. Vila, P. Serrano, K. Wthrich and H.A. Scheraga. Sequential nearest-neighbor effects on computed $^{13}C^{\alpha}$ chemical shifts. J. Biomolec. NMR, 48: 23-30 (2010).

55. 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. Biomolec. NMR, 38: 221-235 (2007).

56. J.A. Vila and H.A. Scheraga. Assessing the accuracy of protein structures by quantum mechanical computations of $^{13}C^{\alpha}$ chemical shifts. Accounts of Chem. Res., 42: 1545-1553 (2009).

57. Y.A. Arnautova, J.A. Vila, O.A. Martin and H.A. Scheraga. What can we learn by computing $^{13}C^{\alpha}$ chemical shifts for X-ray protein models? Acta Cryst. D65: 697-703 (2009).

58. J.A. Vila, Y.A. Arnautova, Y. Vorobjev and H.A. Scheraga. Assessing the fractions of tautomeric forms of the imidazole ring of histidine in proteins as a function of pH. Proc. Natl. Acad. Sci., U.S.A., 108: 5602-5607 (2011).

59. J.A. Vila, Y.A. Arnautova, O.A. Martin and H.A. Scheraga. Quantum-mechanics-derived $^{13}C^{\alpha}$ chemical shift server (*Che*Shift) for protein structure validation. Proc. Natl. Acad. Sci, U.S.A., 106: 16972-16977 (2009).

60. H.A. Scheraga. Respice, Adspice, and Prospice, Ann. Rev. Biophys. 40: 1-39 (2011).

Multilevel Enhanced Sampling of Cellulose-Cellulase Interaction

Emal M. Alekozai[1,2], Jeremy C. Smith[1,2], and Xiaolin Cheng[2]

[1] Interdisciplinary Center for Scientific Computing, University of Heidelberg
Im Neuenheimer Feld 368, 69120 Heidelberg, Germany
E-mail: Emal.Alekozai@iwr.uni-heidelberg.de

[2] Center for Molecular Biophysics, University of Tennessee / Oak Ridge National Laboratory
One Bethel Valley Road, Oak Ridge, TN 37831, USA
E-mail: {AlekozaiEM, SmithJC, ChengX}@ornl.gov

The cellulase enzyme Cel7A is a promising candidate to hydrolyze cellulose efficiently. One of the major research questions is to understand the mechanical process how the cellulase enzyme accesses the cellulose fiber. To guarantee, with the present day computer power, statistically unbiased results a systematic multilevel simulation strategy is required. In the first step multiple Brownian dynamics simulations ($> 3.2\,ms$) are used to study the global cellulose-cellulase interactions, which occur on the microsecond timescale. To refine our results, in the second step multiple Molecular Dynamic simulations ($> 1\,\mu s$) are used to study the local interactions, which occur on the ns timescale. We utilized different methods like density maps and MSM for the analysis. Our study could extend and quantify the many existing experimental results.

1 Introduction

Cellulose holds great potential as a carbon-neutral renewable source of biofuel energy, it is the most abundant bioresource produced in the biosphere ($\approx$ 100 billion dry tons/ year)[1–3]. Cellulose forms the cell walls of plants and gives them rigidity. Cellulose chains are linear polymers consisting of glucose units joined together by β-(1,4)-glycosidic bonds. These single cellulose chains are packed and bundled into ordered arrays to form cellulose fibers. They are held together by van der Waals forces and hydrogen bonds. The glucose units can be unlocked and fermented to produce ethanol. The half life for the $\beta-$ glucosidic bond cleavage at $25°C$ is 5-8 million years, which makes the cellulose molecule very stable. Therefore the fermentation process is difficult[4].

Cellulase enzymes are capable of breaking up the sugar chains. A promising candidate is the cellulase enzyme Cellobiohydrolase (CBH) I, referred to here as Cel7A[5]. It is an industrial important cellulase secreted in high yields by the filamentous fungus *Trichoderma reesei*[6]. It consists of two parts, the carbohydrate binding module (CBM) and the catalytic domain (CD) which are held together by a highly O-glycosylated linker peptide (Fig. 1 top). Cel7A is a processive exocellulase, which prefers to hydrolyze cellulose chains from the reducing to the non-reducing end[7,8]. The CD has an $50\,\mathring{A}$ long tunnel-shaped active site, which can bind and hydrolyze a single cellulose chain. In experiments it is observed that the removal of the CBM reduces the CD activity. The study[9] showed that after the removal of the CBM, Cel7A almost completely lost its ability to bind to cellulose. This implies that the CBM could enhance the cellulase activity by increasing the local enzyme concentration on the cellulose surface[10,11]. The three Tyr residues Y5, Y31 and Y32 form a flat hydrophobic patch on the surface of the CBM, which matches a similar hydrophobic
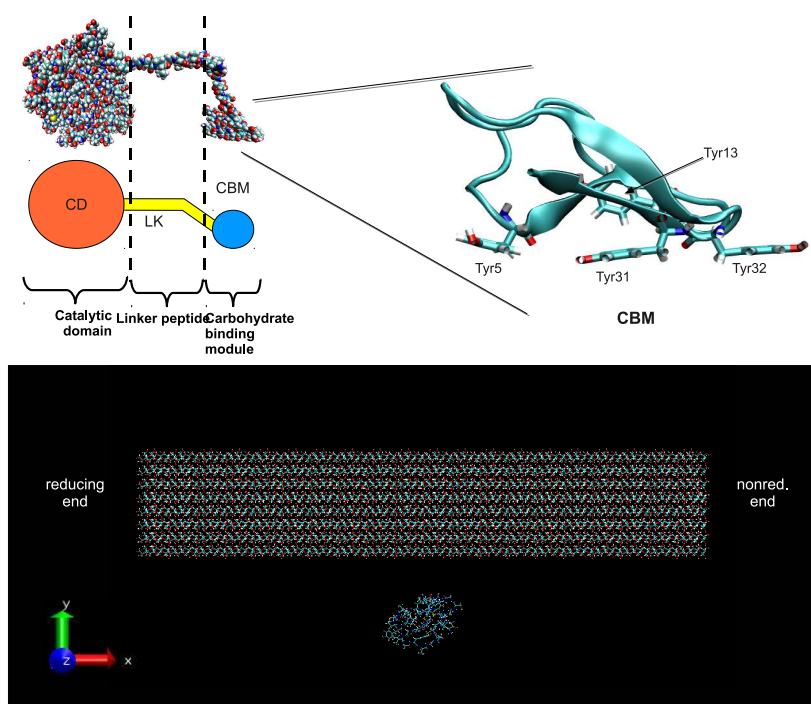
Figure 1. (Top) Visualization of the Cel7A CBM, the residues Y5, Y13, Y31 and Y32 are highlighted. (Bottom) Cel7A CBM together with the cellulose fiber, the largest dimension of the cellulose fiber is oriented parallel to the x-axis of the coordinate system.

surface of the cellulose fiber surface. These three residues may play an important role in the docking of the CBM to the microfibrils by aligning their rings relative to the sugar monomers. Studies[12,13] suggest that the fourth Tyr residue Y13 can move from its internal position to the cellulose surface to form hydrophobic interactions with the cellulose surface. Cellulases are relatively costly enzymes[14], high enzyme loads are currently needed to realize reasonable yields[15,16]. To achieve higher conversion yields a better understanding of the mechanism how the enzyme degenerates the cellulose is important. On the different fiber faces different atoms are pointing out, which lead to different hydrophobicity for the different fiber faces.

To sample the relevant cellulose-cellulase interactions with the present day computer power we choose a multilevel sampling strategy. In the first step we performed multiple Brownian dynamic (BD) simulations, resulting in a simulation time $> 3.2\,ms$, to study the global CBM fiber interactions. We constructed for the CBM-fiber interactions a density map as well as a Markov state model (MSM)[17–21] to identify the fiber "hotspots" with favorable CBM interactions. To refine the results we performed in the second step Molecular dynamics (MD) simulations for the "hotspot" conformations, leading to a combined simulation time of $> 1\,\mu s$.

2 Material and Methods

The CBM has a size of $\approx 29\,\dot{A} \times\, \approx 17\,\dot{A} \times\, \approx 24\,\dot{A}$. It consists of 36 amino acid residues with sequence TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL. A cellulose $I\beta$ fiber model with a length of $\approx 206\,\dot{A}$ and a diameter of $\approx 40\,\dot{A}$ was chosen (Fig. 1 bottom). The model consists of 36 individual chains, each chain consists of 40 glucose units. Based on the ordering of the chains in the end-view eight different fiber faces are visible, which have different physical properties. The fiber faces $(1, 0, 0)$ and $(-1, 0, 0)$ are the most hydrophobic ones and $(0, 1, 0)$ and $(0, -1, 0)$ are the least hydrophobic faces (here referred to as hydrophilic).

In the MD simulations the CBM-fiber system has a size of ≈ 210000 atoms. For the BD simulations the translational D_{trans} and rotational D_{rot} diffusion coefficients were determined from the MD simulations and additionally verified by inputting the structures to the hydrodynamics program HYDROPRO[22]. Prior to analysis the BD trajectories are transformed into a reference system in which the cellulose fiber has no translational or rotational motion, this reduces the number of degrees of freedom of the system from 12 to 6. The coordinate system is chosen such, that the longest fiber dimension is parallel to the x axis. An infinitely long fiber is homogeneous along the x axis, the CBM cannot distinguish between different positions along the x axis, therefore the x axis was projected out. To avoid end effects, only simulation frames are used in which the center of the CBM has a distance of $\epsilon \geq 30\,\dot{A}$ from the fiber ends. The analysis were performed using GNU R[23].
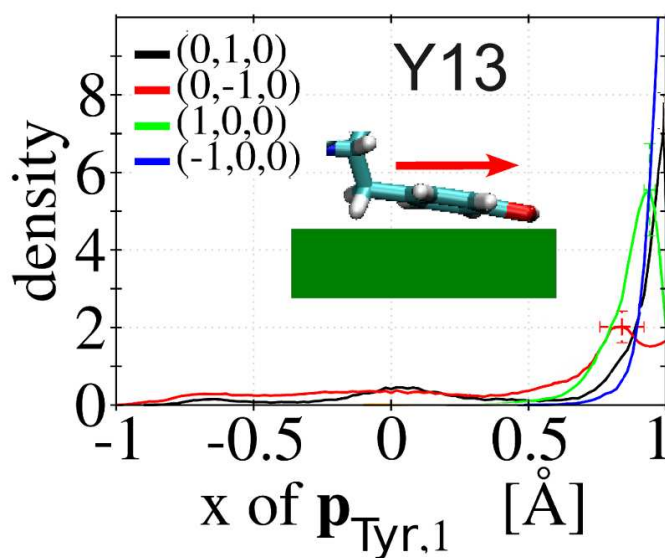
Figure 2. Orientation of the Y13 residue of the Cel7A CBM. The normalized vector $p_{Tyr,1}$ points from the C^γ to the C^ζ atom of the Tyr ring.

3   Results and Discussion

To sample the relevant cellulose-cellulase interactions we used a two step protocol. We utilized BD simulations for the global and MD simulations for the local CBM-fiber interactions. The simulations give new insights on an atomic level with a ps time resolution. This detailed spatial and time resolution are in general not accessible via experiments.The electron microscopy[24], the single molecule fluorescence[25] and the atomic force microscopy[26] studies found that the CBM prefers to bind to the hydrophobic faces of cellulose $I\alpha$. The BD and MD simulations support and extend this observation by giving additional quantitative insights. Here we focus on the orientation of the CBM Tyr residues. For all fiber faces the residues Y31 is pointing towards the reducing end and Y32 is pointing towards the nonreducing end. The residue Y13 is pointing towards the nonreducing end (Fig. 2). The CBM tries by aligning the Tyr residues to expand the hydrophobic surface.

Acknowledgments

References

1. MT Holtzapple, *Cellulose In, Macrae R, Robinson RK, Saddler MJ, editors. Encyclopedia of food science food technology and nutrition*, London: Academic Press, **16**, 758–767, 1993.

2. M. Jarvis, *Cellulose stacks up*, Nature, **426**, no. 6967, 611–612, 2003.

3. Yi-Heng Percival Zhang and Lee R Lynd, *Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems.*, Biotechnol Bioeng, **88**, no. 7, 797–824, Dec 2004.

4. M.E. Himmel, S.Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, and T.D. Foust, *Biomass recalcitrance: engineering plants and enzymes for biofuels production*, Science, **315**, no. 5813, 804, 2007.

5. L. Zhong, J.F. Matthews, M.F. Crowley, T. Rignall, C. Talón, J.M. Cleary, R.C. Walker, G. Chukkapalli, C. McCabe, M.R. Nimlos, et al., *Interactions of the complete cellobiohydrolase I from Trichodera reesei with microcrystalline cellulose Iβ*, Cellulose, **15**, no. 2, 261–273, 2008, Cellulose potential energy is stored in its C-H and C-C bonds.

6. M. J. Harrison, A. S. Nouwens, D. R. Jardine, N. E. Zachara, A. A. Gooley, H. Nevalainen, and N. H. Packer, *Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of Trichoderma reesei.*, Eur J Biochem, **256**, no. 1, 119–127, Aug 1998.

7. Mária Vrsanská and Peter Biely, *The cellobiohydrolase I from Trichoderma reesei QM 9414: action on cello-oligosaccharides*, Carbohydrate Research, **227**, 19 – 27, 1992.

8. Tuula T. Teeri, *Crystalline cellulose degradation: new insight into the function of cellobiohydrolases*, Trends in Biotechnology, **15**, no. 5, 160 – 167, 1997, Reference for exo- and endcellulase.

9. AV Gusakov, AP Sinitsyn, AV Markov, AA Skomarovsky, OA Sinitsyna, AG Berlin, and NV Ankudimova, *Indigo-binding domains in cellulase molecules*, Biocatalysis-2000: fundamentals and applications, **41**, no. 6, 77–80, 2000.

10. J. Knowles, P. Lehtovaara, and T. Teeri, *Cellulase families and their genes*, Trends in Biotechnology, **5**, no. 9, 255–261, 1987.

11. T.T. Teeri, T. Reinikainen, L. Ruohonen, T.A. Jones, and J.K.C. Knowles, *Domain function in Trichoderma reesei cellobiohydrolases*, Journal of Biotechnology, **24**, no. 2, 169–176, 1992.

12. Lintao Bu, Gregg T Beckham, Michael F Crowley, Christopher H Chang, James F Matthews, Yannick J Bomble, William S Adney, Michael E Himmel, and Mark R Nimlos, *The energy landscape for the interaction of the family 1 carbohydrate-binding module and the cellulose surface is altered by hydrolyzed glycosidic bonds.*, J Phys Chem B, **113**, no. 31, 10994–11002, Aug 2009.

13. M.R. Nimlos, J.F. Matthews, M.F. Crowley, R.C. Walker, G. Chukkapalli, J.W. Brady, W.S. Adney, J.M. Cleary, L. Zhong, and M.E. Himmel, *Molecular modeling suggests induced fit of Family I carbohydrate-binding modules with a broken-chain cellulose surface*, Protein Engineering Design and Selection, **20**, no. 4, 179, 2007.

14. Y.H. Percival Zhang, M.E. Himmel, and J.R. Mielenz, *Outlook for cellulase improvement: screening and selection strategies*, Biotechnology Advances, **24**, no. 5, 452–481, 2006.

15. M.E. Himmel, M.F. Ruth, and C.E. Wyman, *Cellulase for commodity products from cellulosic biomass*, Current opinion in biotechnology, **10**, no. 4, 358–364, 1999.

16. R. Wooley, M. Ruth, D. Glassner, and J. Sheehan, *Process design and costing of bioethanol technology: a tool for determining the status and direction of research and development*, Biotechnology Progress, **15**, no. 5, 794–803, 1999.

17. P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation and queues*, Springer, 1999.

18. Nina Singhal, Christopher D. Snow, and Vijay S. Pande, *Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin*, The Journal of Chemical Physics, **121**, no. 1, 415–425, 2004.

19. Nina Singhal and Vijay S. Pande, *Error analysis and efficient sampling in Markovian state models for molecular dynamics*, The Journal of Chemical Physics, **123**, no. 20, 204909, 2005.

20. N.S. Hinrichs and V.S. Pande, *Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics*, The Journal of Chemical Physics, **126**, 244101, 2007.

21. S.P. Elmer, S. Park, and V.S. Pande, *Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular-dynamics simulations in acetonitrile, chloroform, methanol, and water*, The Journal of Chemical Physics, **123**, 114902, 2005.

22. J. García de la Torre, M.L. Huertas, and B. Carrasco, *Calculation of hydrodynamic properties of globular proteins from their atomic-level structure*, Biophysical journal, **78**, no. 2, 719–730, 2000.

23. R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN 3-900051-07-0.

24. Janne Lehtiö, Junji Sugiyama, Malin Gustavsson, Linda Fransson, Markus Linder, and Tuula T. Teeri, *The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules*, Proceedings of the National Academy of Sciences of the United States of America, **100**, no. 2, 484–489, 2003.

25. D.J. Dagel, Y.S. Liu, L. Zhong, Y. Luo, M.E. Himmel, Q. Xu, Y. Zeng, S.Y. Ding, and S. Smith, *In Situ Imaging of Single Carbohydrate-Binding Modules on Cellulose Microfibrils*, The Journal of Physical Chemistry B, **115**, 635–641, 2010.

26. Y.S. Liu, J.O. Baker, Y. Zeng, M.E. Himmel, T. Haas, and S.Y. Ding, *Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces*, Journal of Biological Chemistry, **286**, no. 13, 11195, 2011.

Effects of Confinement on the Thermodynamics of a Model Protein

Mustafa Bilsel[1], Buket Taşdizen[1], Handan Arkin[1,2], and Wolfhard Janke[2]

[1] Department of Physics Engineering, Faculty of Engineering, Ankara University
Tandoğan, Ankara, Turkey

[2] Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, 04009 Leipzig, Germany
E-mail: Handan.Arkin@itp.uni-leipzig.de

We report the results of computer simulations of a model protein confined in a cage like a sphere to investigate the dynamics of the folding mechanism. The problem of whether proteins are misfolded or aggregated or on the contrary are folded properly more promptly in these confining media is of great interest in our study. Therefore our aim is to analyze the thermodynamics of the folding mechanism and also to investigate whether the folding mechanism is controlled or not in the confining media. To do so we have employed exhaustive multicanonical Monte Carlo simulations by using a minimalistic AB model approach. A detailed understanding of this subject plays a key role for finding treatments to diseases caused by misfolding of proteins.

1 Introduction

Protein folding is one of the most intensively studied and still unsolved problems in biology. The process by which a protein folds into its biologically active state cannot be traced in all details solely by experiments. Therefore, many theoretical and experimental studies focus on determination of the three-dimensional structure of these molecules. Recently, molecular modelling has attracted considerable attention for applications in designing and fabrication of nanostructures leading to the development of advanced materials. In a newly growing field of research, synthetic peptides are investigated for their use in nano-devices, by exploiting their self-assembly properties[1,2]. The self-assembly of biomolecular building blocks plays an increasingly important role in the discovery of new materials, with a wide range of applications in nanotechnology and medical technologies such as drug delivery systems[3]. In these studies, several types of biomaterials are developed, ranging from models for studying protein folding to molecular materials for producing peptide nanofibers, peptide surfactants by designing various classes of self-assembling peptides[4]. These experiments reveal many different interesting and important problems, which are related to general aspects of the question why and how proteins fold. In this context, modern simulation techniques have opened another window to give a new insight to the protein folding problem[5].

In this study, we focus on the folding of the model protein $BA_6BA_4BA_2BA_2B_2$ under the influence of a confining potential which simulates a cage being composed of rigid walls.

2 Model

The polymer chains are described by a coarse-grained hydrophobic-polar model which also helped to understand protein folding channels from a mesoscopic perspective[6]. A manifest

off-lattice variant of the HP model[7] is the AB model[8], where the hydrophobic monomers are labeled by A and the polar or hydrophilic ones by B. As on the lattice, the adjacent monomers are connected by rigid covalent bonds. Thus, the distance is fixed and set to unity. The contact interaction is replaced by a distance-dependent Lennard-Jones type of potential accounting for short-range excluded volume repulsion and long-range interaction. An additional interaction accounts for the bending energy of any pair of successive bonds. This model was first applied in two dimensions[8] and generalized to three-dimensional AB proteins[9], partially with modifications taking implicitly into account additional torsional energy contributions of each bond.

The AB model as proposed in Ref. 9 has an energy function

$$E = -\kappa_1 \sum_{k=1}^{N-2} b_k . b_{k+1} - \kappa_2 \sum_{k=1}^{N-3} b_k . b_{k+2} + 4 \sum_{i=1}^{N-2} \sum_{j=i+1}^{N} C(\sigma_i, \sigma_j) \left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}} \right) \quad (1)$$

where b_k is the bond vector between the monomers k and $k+1$ with length unity. The second term in Eq. (1) takes torsional interactions into account without being an energy associated with the pure torsional barriers in the usual sense. The third term contains now a pure Lennard-Jones potential, where the $1/r_{ij}^6$ long-range interaction is attractive whatever types of monomers interact. The monomer-specific prefactor $C(\sigma_i, \sigma_j)$ only controls the depth of the Lennard-Jones valley:

$$C(\sigma_i, \sigma_j) = \begin{cases} +1, & \sigma_i, \sigma_j = A, \\ +1/2, & \sigma_i, \sigma_j = B \quad \text{or} \quad \sigma_i \neq \sigma_j. \end{cases} \quad (2)$$

Simulations of this model were performed with the multicanonical algorithm[10] and the update mechanism is a spherical update which is described in Ref. 11 in detail.

3 Confining Potentials

The focus of this study is to comprehend the folding mechanism of proteins in their cellular environments. To emulate this effect, the following potentials are used:

$$V_1(r) = \frac{0.01}{R_c} \left[e^{r-R_c}(r-1) - \frac{r^2}{2} \right], \quad (3)$$

$$V_2(r) = 4\epsilon_c \frac{\pi R_c}{r} \left(\frac{1}{5} \left[\left(\frac{\sigma}{r - R_c} \right)^{10} - \left(\frac{\sigma}{r + R_c} \right)^{10} \right] \right), \quad (4)$$

$$V_3(r) = 4\epsilon_c \frac{\pi R_c}{r} \left(\frac{1}{5} \left[\left(\frac{\sigma}{r - R_c} \right)^{10} - \left(\frac{\sigma}{r + R_c} \right)^{10} \right] \right. $$
$$\left. - \frac{\epsilon}{2} \left[\left(\frac{\sigma}{r - R_c} \right)^{4} - \left(\frac{\sigma}{r + R_c} \right)^{4} \right] \right) \quad (5)$$

where R_c is the sphere radius which is a measure of the cage size, $r = (x^2 + y^2 + z^2)^{1/2}$ is the distance of a monomer to the origin and x, y, z are the coordinates of monomers, $\sigma = 1.0$ and $\epsilon_c = \epsilon = 1.0$. For our simulations, we set R_c large enough to enclose the protein inside the sphere.
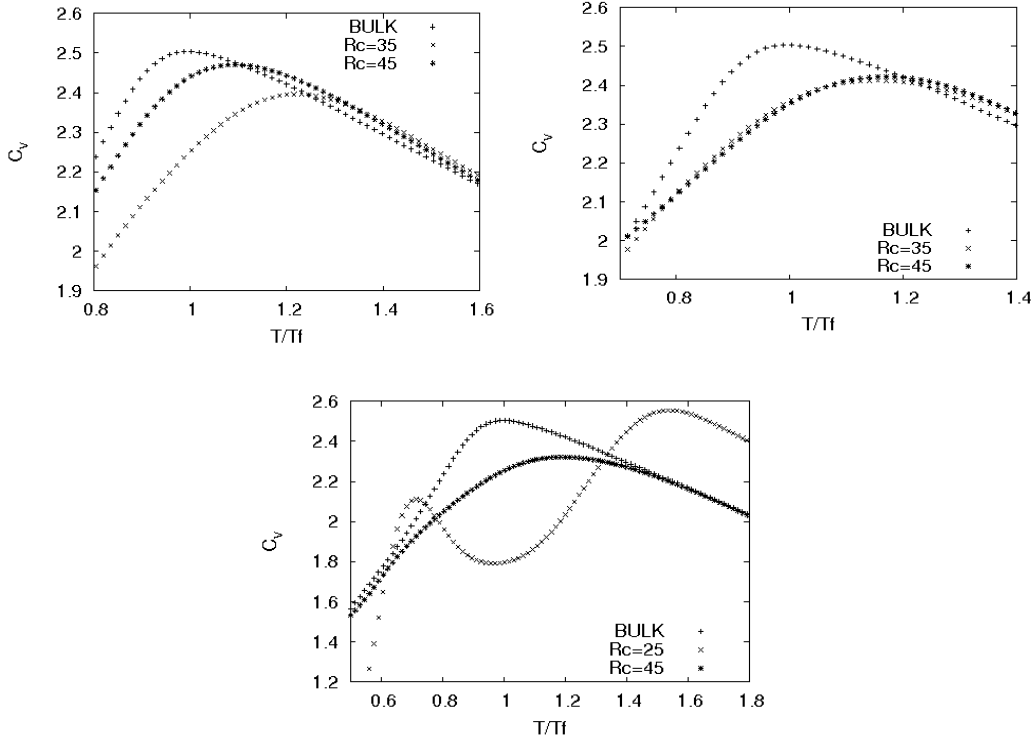
Figure 1. Specific-heat plots under the influence of potential a) V_1, b) V_2, and c) V_3.

4 Results and Discussion

The effect of confinement on the thermodynamic properties of the model protein was investigated by multicanonical simulations. We have chosen appropriate radii such that proteins are not allowed to move outside the sphere. The results are compared to a bulk protein which folds at $T = T_f$. After calculating multicanonical weights, 5×10^7 iterations were performed in the production run. We computed the transition temperature for the bulk protein and protein in cages with different radii. The results are plotted in Fig. 1. When the radius increases the effect of the confining potential decreases so that $R_c = 100$ behaves like a bulk environment.

The specific heats of potential V_1, V_2 and V_3 as a function of temperature are plotted in Fig. 1a), b) and c), respectively. Generically, increase in the radius of the sphere causes a decrease of the transition temperature. On the other hand, the situation for V_3 is different. The V_3 potential contains an attractive part so the model protein feels the effect of attractiveness of the potential at small radius ($R_c = 25$), and is adsorbed by the surface of the sphere in a first stage and in a second stage it arranges its structure. By increasing the radius up to $R_c = 45$ the influence of the potential starts to decrease and so adsorption disappears.

As a result, for all cases specific-heat plots were broadened when the radius of the sphere decreases. Narrower graph means states close to native ones appear more often than in the bulk situation. We can conclude that the protein becomes more stable when the radius of the sphere decreases. These results are compatible with the previous ones[12–14].

Acknowledgments

References

1. S. Santosa, W. Hwang, H. Hartman, and S. Zhang, *Self-asembly of surfactant-like peptides with variable glycine tails to form nanotubes and nanovesicles*, Nano Lett. **2**, 687–691, 2002.
2. S. Vauthey, S. Santosa, H. Gong, N. Watson, and S. Zhang, *Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles*, Proc. Nat. Acad. Sci. USA **99**, 5355–5360, 2002.
3. J. A. Hubbell, *Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering*, Curr. Opin. Biotechnol. **10**, 123–128, 1999.
4. T. Holmes, S. De Lacalle, X. Su, A. Rich, and S. Zhang, *Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds*, Proc. Natl. Acad. Sci. USA **97**, 6728–6733, 2000.
5. H. Arkin, *Searching low energy conformations of two elastin sequences*, Eur. Phys. J. B. **37**, 223–228, 2004.
6. S. Schnabel, M. Bachmann, and W. Janke, *Two-state folding, folding through intermediates, and metastability in a minimalistic hydrophobic-polar model for proteins*, Phys. Rev. Lett **98**, 048103-1–4, 2007. *Identification of characteristic protein folding channels in a coarse-grained hydrophobic-polar peptide model*, J. Chem. Phys. **126**, 105102-1–6, 2007.
7. K. A. Dill, *Theory for the folding and stability of globular proteins*, Biochemistry **25**, 1501–1509, 1985.
8. F. H. Stillinger and T. Head-Gordon, *Collective aspects of protein folding illustrated by a toy model*, Phys. Rev. E **52**, 2872–2877, 1995.
9. A. Irbäck, C. Peterson, F. Potthast, and O. Sommelius, *Monte Carlo procedure for protein design*, Phys. Rev. E **58**, R5249–R5252, 1998.
10. B. A. Berg and T. Neuhaus, *Multicanonical ensemble: A new approach to simulate first-order phase transitions*, Phys. Rev. Lett. **68**, 9–12, 1992.
11. M. Bachmann, H. Arkin, and W. Janke, *Multicanonical study of coarse-grained off-lattice models for folding heteropolymers*, Phys. Rev. E **71**, 031906-1–11, 2005.
12. 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**, 11367–11372, 2003.
13. N. Rathore, T. A. Knotts, and J. J. de Pablo, *Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations*, Biophys. J. **90**, 1767–1773, 2006.
14. D. Lu, Z. Liu, and J. Wu, *Structural transitions of confined model proteins: molecular dynamics simulation and experimental validation*, Biophys. J. **90**, 3224–3238, 2006.

Global Dynamics of Protein and DNA in Nucleosome Investigated by Molecular Dynamics Simulation

Mithun Biswas[1]**, Jörg Langowski**[2]**, and Jeremy C. Smith**[3]

[1] Computational Molecular Biophysics, IWR, University of Heidelberg
Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany
E-mail: mithun.biswas@iwr.uni-heidelberg.de

[2] Biophysics of Macromolecules, German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
E-mail: joerg.langowski@dkfz-heidelberg.de

[3] University of Tennessee / Oak Ridge National Laboratory
Center for Molecular Biophysics, Oak Ridge National Laboratory
P.O. Box 2008, Oak Ridge, Tennessee 37831-6164, USA
E-mail: smithjc@ornl.gov

Nucleosomes are the basic compacting units of eukaryotic chromatin in which DNA is tightly wrapped around histone proteins. Structural transitions in nucleosome open up DNA surface to regulatory factors. Here we report a principal component analysis (PCA) performed on a 100 ns nucleosome trajectory to obtain insights into the dominant motions of protein and DNA in nucleosome.

1 Introduction

Eukaryotic DNA is hierarchically organized into chromatin around single repeating units known as nucleosomes. In a nucleosome about 150 bp of DNA are wrapped in left-handed superhelical turns around an octameric histone protein complex. The histone octamer has a tripartite structure composed of a (H3-H4)$_2$ tetramer flanked by two H2A-H2B dimers.

The 1.9 Å resolution structure of the nucleosome core particle revealed interactions between the histone core, histone tails and DNA at atomic detail[1]. In the structure the four histone dimers (two each of H3, H4, H2A and H2B) are arranged about a two-fold dyad symmetry axis, which also intersects with the middle of the DNA fragment (Fig. 1). The histone proteins are composed of structured core domains, mainly α-helices, and unstructured tails passing between the DNA superhelix turns. The DNA is wound tightly around the histone protein and held in place by 14 histone-DNA contact regions.

The nucleosomal organization of eukaryotes provides protection against possible DNA damage and regulates access to it. Evidences suggest that during different periods of cell cycle DNA access is regained by a number of different mechanisms : the transient dissociation and rebinding of peripheral DNA from nucleosome, enzyme catalyzed or thermally driven nucleosome sliding, incorporation of histone variants to nucleosome having higher mobility compared to the wild-type and histone post-translational modifications[2].

Spontaneous structural changes in the nucleosome is one of the primary mechanisms that opens up the DNA surface to regulatory factors. One way to identify the structural states of nucleosome is to investigate large amplitude collective modes of nucleosome dynamics. In the present proceedings we report a principal component analysis (PCA)
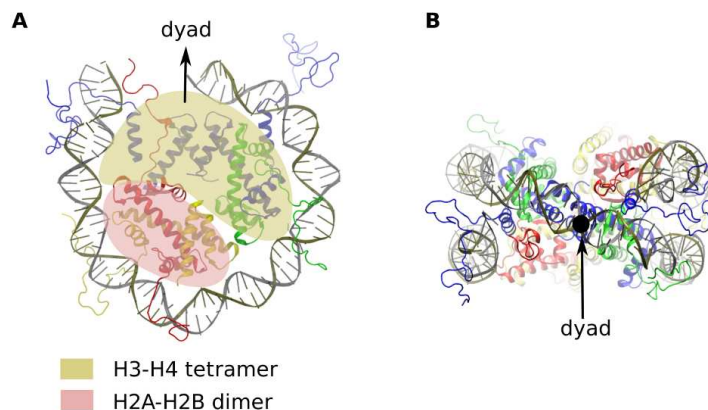
Figure 1. Structure of the nucleosome (1kx5). (A) Half of the nucleosomal DNA with locations of the (H3-H4)$_2$ tetramer and H2A-H2B dimer. The dyad axis is indicated by arrow. (B) Side view of the nucleosome with the dyad axis.

performed on 100 ns molecular dynamics (MD) trajectory enabling identification of slow collective motions of the nucleosome.

2   Methods

The starting structure of all molecular dynamics (MD) simulations was taken from the 1.9 Å resolution crystal structure of the nucleosome core particle (PDB ID: 1KX5)[1]. All simulations used the CHARMM27 force field[3] in the NAMD program[4]. The structures were immersed in a cubic box of TIP3P water molecules and there was at least 10 Å of separation between the solute and the edge of the box. The system was then neutralized with Na$^+$ ions and appropriate amount of NaCl was added to keep the systems at 150 mM salt concentration. Periodic boundary conditions were used and the long range electrostatics was treated with the particle mesh Ewald method[5]. For the van der Waals interactions a switching function was applied at 10 Å and the cut-off was set to 12 Å. The integration time step was 2 fs. The pressure was kept constant at atmospheric pressure at sea level with the Nos-Hoover Langevin piston pressure control[6,7] in NAMD. The temperature was maintained at 300K with a Langevin damping coefficient of 2 ps^{-1}.

Given the $3N$ cartesian atomic coordinates $x_i(t)$ $(i = 1, 2, ..., 3N)$ the elements of the covariance matrix C of atomic positions are given as $C_{ij} = \langle (x - \langle x_i \rangle)(y - \langle y_j \rangle) \rangle$. Diagonalization of C yields $3N$ orthonormal eigenvectors, also called principal component vectors, with variance σ_i^2 as corresponding eigenvalues, ordered in descending magnitude. Principal components are projections of the atomic trajectory onto principal component vectors. The first few principal components represent large amplitude collective motion of the system along independent directions.

3 Results

To investigate collective modes of nucleosome dynamics the motion of the (H3-H4)$_2$ tetramer, two H2A-H2B dimers and the DNA during the 100 ns trajectory were analyzed separately using PCA of Cα atoms for the histones and phosphorous atoms for the DNA. The analysis of eigenvalues for histones and DNA indicate that the first two eigenmodes account for $\sim 30\%$ of the total atomic fluctuation.
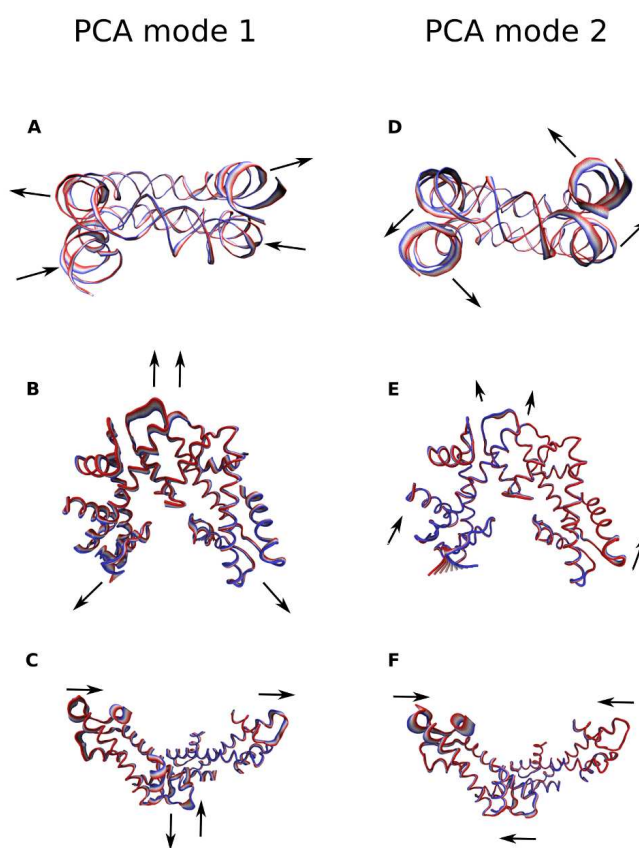
PCA mode 1   PCA mode 2

Figure 2. Global motion of histone (excluding histone tails) and DNA in the nucleosome. (A)-(B)-(C) Dynamics of the nucleosomal DNA, (H3-H4)$_2$ tetramer and the H2A-H2B dimers respectively in PCA mode 1. The deformation is illustrated by superposition of structures with the extremes indicated by blue and red tubes. Approximate directions of motion are indicated by arrows. (D)-(E)-(F) Dynamics of the histone domains and the DNA in PCA mode 2.

In nucleosome the DNA is wrapped around histones in about two rings. In the first mode the two DNA rings undergo anti-correlated compression and relaxation (Fig. 2A). This mode of DNA dynamics would support the propagation of a DNA bulge along nucleosomal DNA which has been proposed as one of the key mechanisms of nucleosome

sliding[8]. Concurrently, the (H3-H4)$_2$ tetramer gets compressed and relaxed along the dyad axis (Fig. 2B). Unlike the (H3-H4)$_2$ tetramer which contacts the DNA near the dyad axis, the two H2A-H2B dimers lie in two separate planes defined by the DNA rings and, similar to the motion of DNA in the two rings, the H2A-H2B dimers undergo anti-correlated motion (Fig. 2C).

In the second mode the DNA, the (H3-H4)$_2$ tetramer and the H2A-H2B dimers tends to bend with respect to the dyad axis in an out-of-plane fashion (Fig. 2D, 2E, 2F). The out-of-plane DNA motion may aid dissociation of peripheral regions of DNA from the histone octamer which has been observed in experiments[9]. In transcriptionally active chromatin canonical histone H2A is rapidly exchanged with the H2A variants to regulate genomic access[2]. The out-of-plane anti-correlated distortions of H2A-H2B dimers along the second PCA mode is likely indicative of the motion associated with the disassembly of histone H2A from the nucleosome.

Acknowledgments

References

1. Curt A Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond, *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution.*, J Mol Biol, **319**, no. 5, 1097–1113, Jun 2002.
2. Karolin Luger, *Dynamic nucleosomes.*, Chromosome Res, **14**, no. 1, 5–16, 2006.
3. B. R. Brooks et. al., *CHARMM: the biomolecular simulation program.*, J Comput Chem, **30**, no. 10, 1545–1614, Jul 2009.
4. James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kal, and Klaus Schulten, *Scalable molecular dynamics with NAMD.*, J Comput Chem, **26**, no. 16, 1781–1802, Dec 2005.
5. Tom Darden, Darrin York, and Lee Pedersen, *Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems*, The Journal of Chemical Physics, **98**, no. 12, 10089–10092, 1993.
6. Glenn Martyna, Douglas Tobias, and Michael Klein, *Constant pressure molecular dynamics algorithms*, The Journal of Chemical Physics, **101**, no. 5, 4177–4189, 1994.
7. Scott Feller, Yuhong Zhang, Richard Pastor, and Bernard Brooks, *Constant pressure molecular dynamics simulation: The Langevin piston method*, The Journal of Chemical Physics, **103**, no. 11, 4613–4621, 1995.
8. H. Schiessel, J. Widom, R. F. Bruinsma, and W. M. Gelbart, *Polymer Reptation and Nucleosome Repositioning*, Phys. Rev. Lett., **86**, no. 19, 4414–, May 2001.
9. K. J. Polach and J. Widom, *Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation.*, J Mol Biol, **254**, no. 2, 130–149, Nov 1995.

Protein Structure Prediction Using CABS
– A Consensus Approach

Maciej Blaszczyk, Michal Jamroz, Dominik Gront, and Andrzej Kolinski

Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw
02-093 Warsaw, Poland
E-mail: mblaszczyk@chem.uw.edu.pl

We have designed a new pipeline for protein structure prediction based on the CABS engine. The procedure is fully automated and generates consensus models from a set of templates. Restraints derived from the templates define a region of conformational space, which is then sampled by Replica Exchange Monte Carlo algorithm implemented in CABS. Results from CASP9 show, that for great majority of targets this approach leads to better models than the mean quality of templates (in respect to GDT_TS). In five cases the obtained models were the best among all predictions submitted to CASP9 as the first models.

1 Introduction

Knowledge of 3D structures of proteins is a crucial requirement for a progress in many areas of biomedicine, e.g. rational drug design. Due to the complexity and high cost of structure determination by experimental methods (mainly Xray crystallography or NMR), computer-based protein structure prediction methods have been placed in the center of attention of a broad community of molecular and cell biologists[1]. Nowadays, there is a number of publicly available web servers, which provide methods for protein structure prediction[2]. Moreover, thanks to the meta-servers[3,4], which collect data from servers, obtaining the predictions is even easier. However, for most purposes it is necessary to provide one, possibly the best, final model. A common approach to this problem is the use of Model Quality Assessment Programs (MQAPs) which score models according to various criteria[5] and selection of the top scoring one. Obviously, the MQAPs can't propose a model better then the best of input structures. Application of CABS modeling tool[6] with spatial restraints derived from the templates allows for reaching beyond this limit.

2 Methods

The procedure used during CASP9 consisted of several steps (Fig. 1) and was trained on the targets from previous CASPs. The first step was templates selection. As templates we used server predictions submitted to CASP9. The list of the servers from which models were taken, was created on the basis of their performance during the CASP8. To check if the best servers from CASP8 are still the reliable ones, servers predictions from CASP9 were ranked using 3D-jury score[7]. Then, for all selected templates distances between pairs of alpha carbons were extracted[8]. The minimum and the maximum distance between pairs of residues were taken as limits of the ranges of restraints. Using templates as a starting structures we have run two independent Replica Exchange Monte Carlo simulations with CABS[6].
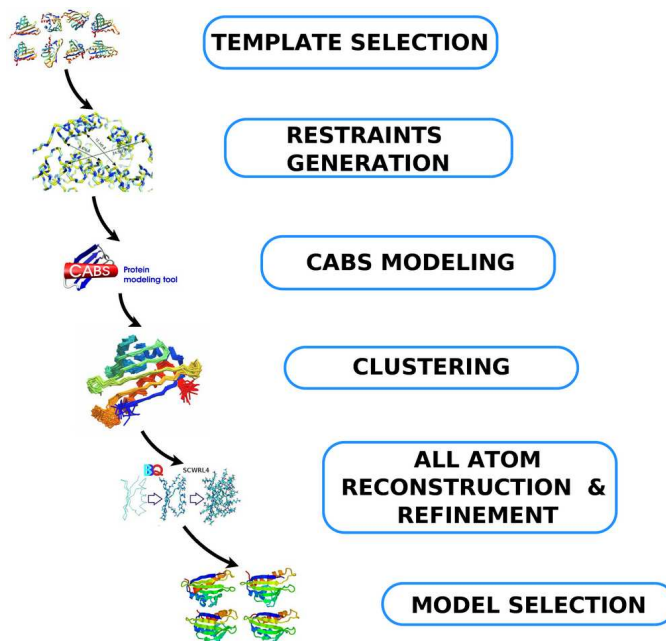
Figure 1. Flowchart of the pipeline used during CASP9. See the text for details.

CABS is a lattice model with a representation reduced to four united atoms per residue: $C\alpha$, $C\beta$, center of mass of a side chain (where applicable) and the center of a virtual $C\alpha -$ $C\alpha$ bond. The force field of the model employs knowledge based potentials derived from the statistical analysis of the databases containing known protein structures. Conformational space is sampled using Replica Exchange Monte Carlo method. Application of the restraints reduces conformational space for sampling, which makes modeling faster and more accurate.

The resulted trajectories from CABS were clustered[9], and the clusters' centroids were calculated. Because of reduced representation in CABS, it was necessary to rebuilt the atomistic details of obtained models. Reconstruction of the backbone using BBQ[10] was followed by reconstruction of the side chains with SCWRL4[11]. Next, we performed model refinement, which was also done in two steps. To improve model geometry (e.g. bond length) we employed Modeller[12]. Then, we used GROMACS[13] in order to refine some packing details. Finally, obtained models were ranked on the basis of the clusters' density and the level of similarity of the models from two independent simulations.

3   Results

Since the presented method aims at a consensus prediction from a set of templates it is worth to compare the accuracy of obtained models and the templates used. For great majority of targets GDT_TS of the model was higher then mean GDT_TS of templates. Moreover, in 5 cases the accuracy of the model was better then the accuracy of the best template (see Fig. 2).
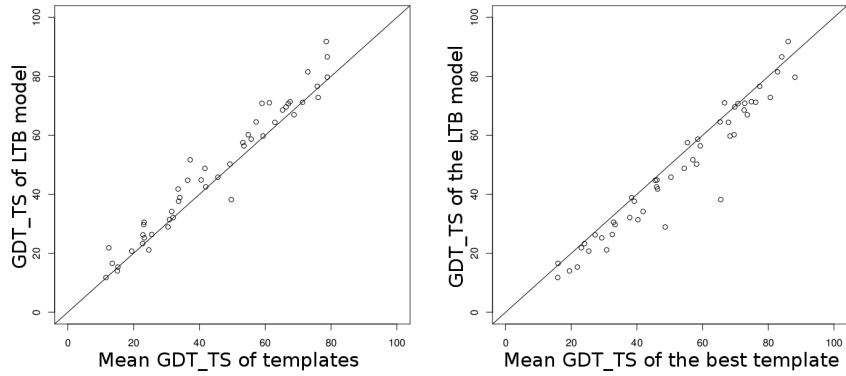
Figure 2. Comparison of GDT_TS of templates and obtained models.

According to the official assessment our models (from Laboratory of Theory of Biopolymers - LTB) for 5 selected domains were the best among all predictions submitted to CASP9 as the first models. As shown in Fig. 3, for great majority of targets, GDT_TS of obtained structure was higher then mean GDT_TS of all models submitted to the CASP. However, there are a few cases with significant losses of accuracy . Most of them are large multi-domain proteins, for which it was necessary to perform domain division, which was not supported in the procedure. This problem is to be solved in a future work.
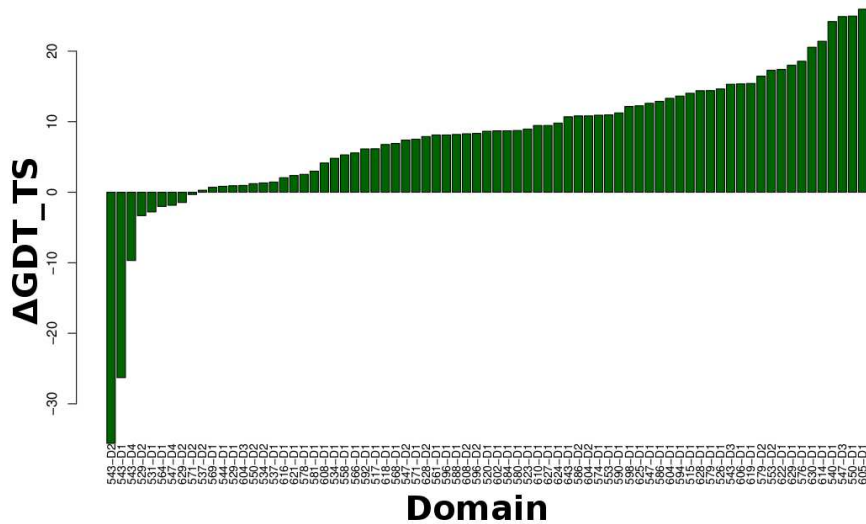
Figure 3. Differences between GDT_TS scores of our models and the mean for all models submitted to CASP.

Acknowledgments

References

1. Yang Zhang, *Protein structure prediction: when is it useful?*, Current opinion in structural biology, **19**, no. 2, 145–155, Apr. 2009.
2. Daniel Fischer, *Servers for protein structure prediction.*, Current opinion in structural biology, Mar. 2006.
3. Krzysztof Ginalski, Arne Elofsson, Daniel Fischer, and Leszek Rychlewski, *3D-Jury: a simple approach to improve protein structure predictions*, Bioinformatics, **19**, no. 8, 1015–1018, May 2003.
4. Jesper Lundström, Leszek Rychlewski, Janusz Bujnicki, and Arne Elofsson, *Pcons: a neural-network-based consensus predictor that improves fold recognition.*, Protein science : a publication of the Protein Society, **10**, no. 11, 2354–2362, Nov. 2001.
5. Andriy Kryshtafovych and Krzysztof Fidelis, *Protein structure prediction and model quality assessment.*, Drug discovery today, **14**, no. 7-8, 386–393, Apr. 2009.
6. Andrzej Kolinski, *Protein modeling and structure prediction with a reduced representation.*, Acta biochimica Polonica, **51**, no. 2, 349–371, 2004.
7. László Kaján and Leszek Rychlewski, *Evaluation of 3D-Jury on CASP7 models*, BMC Bioinformatics, **8**, 304+, Aug. 2007.
8. Dominik Gront and Andrzej Kolinski, *Utility library for structural bioinformatics*, Bioinformatics, **24**, no. 4, 584–585, Feb. 2008.
9. Dominik Gront and Andrzej Kolinski, *HCPM–program for hierarchical clustering of protein models.*, Bioinformatics, **21**, no. 14, 3179–3180, July 2005.
10. Dominik Gront, Sebastian Kmiecik, and Andrzej Kolinski, *Backbone building from quadrilaterals: A fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates*, J. Comput. Chem., **28**, no. 9, 1593–1597, July 2007.
11. Adrian A. Canutescu, Andrew A. Shelenkov, and Roland L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction*, Protein Science, **12**, no. 9, 2001–2014, Sept. 2003.
12. Narayanan Eswar, Ben Webb, Marc A. Marti-Renom, M. S. Madhusudhan, David Eramian, Min-yi Shen, Ursula Pieper, and Andrej Sali, *Comparative Protein Structure Modeling Using Modeller*, Current protocols in bioinformatics, **Chapter 5**, Oct. 2002.
13. David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. Berendsen, *GROMACS: fast, flexible, and free.*, Journal of computational chemistry, **26**, no. 16, 1701–1718, Dec. 2005.

Parallelization of an Efficient Method for Calculating Born Radii

Martin Brieg[1] **and Wolfgang Wenzel**[2]

[1] Steinbuch Centre for Computing, Karlsruhe Institute of Technology
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
E-mail: martin.brieg@kit.edu

[2] Institute of Nanotechnology, Karlsruhe Institute of Technology
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

The Generalized Born Model is one of the most widely used implicit solvent model for simulating biomolecules. Its accuracy and speed crucially depend on the correct calculation of the Born radii used for estimation of the polar solvation free energy. While many improvements in accuracy have been reported over the last years, their implementation in an efficient method remained challenging. Here we present results for the parallelization of a novel method for computing these Born radii on modern multi-core shared memory architectures. This development should significantly improve the acceptance and capabilities of the implicit Generalized Born solvent model.

1 Introduction

The development of efficient yet accurate implicit representations of the environment remains one of the important computational challenges for biomolecular simulation. While numerical solutions to the Poisson Boltzmann (PB) equation provide an accurate estimate of these effects[1], their application in long Molecular Dynamics or Monte Carlo simulations is very time consuming[2]. The most widely accepted approximation to PB calculations is the Generalized Born (GB) Model[3], which estimates the polar solvation free energy by the formula:

$$\Delta G_{GB} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_W}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j}} \tag{1}$$

The accuracy of this estimate crucially depends on the correct calculation of the Born radii R_i which describe the burial of an atom in the solute molecule[4]. Circumventing expensive PB calculations, Grycuk[5] showed that for a spherical solute these are given by the following integral over the solvent region, which also yields reliable results for most non-spherical shapes[6]

$$\frac{1}{R_i^3} = \frac{3}{4\pi}\int_{solvent}\frac{d^3\mathbf{r}}{|\mathbf{r} - \mathbf{r_i}|^6}. \tag{2}$$

Solving these integrals poses the main computational cost within GB models. Though many methods have been developed to tackle this problem, analytical models being limited to spherical surfaces[7–12] require costly correction terms for small interstitial cavities while numerical algorithms are computationally more expensive due to the integration procedures or expensive surface triangulation[13,14]. Here we present results on the parallelization

33
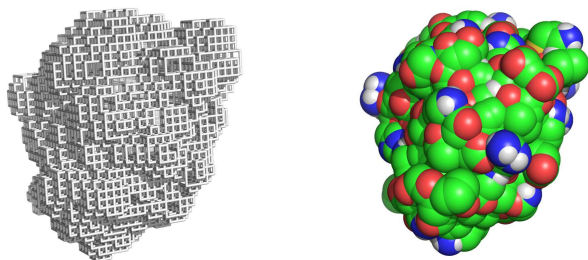
Figure 1. Octree representation of PDB 1OCA (left) and the according solvent accessible surface (right).

of an efficient algorithm, called PowerBornRadii, to compute these integrals for arbitrary surface definitions.

2 Methods

For the algorithm to be independent on the surface definition, it is split into 2 steps as reported elsewhere[15]. The first one consists of constructing a bounding cubic box around the molecule and creating an octree representation of the solute and solvent volume inside this box, see Fig. 1, depending on the chosen surface definition.

In the second step the integral outside the bounding box is evaluated analytically by converting the volume integral to a surface integral and solving it for all six sides of the cube. The integration over the bounding box's inside is computed numerically by integrating over the octree cells not belonging to the solute volume. This is either accomplished by employing the analytical formulas for the integration over the outside of bounding box with an inverse sign or with a fitted approximate formula depending on the size and distance of the octree cell to the location $\mathbf{r_i}$ of the atom of interest.

3 Results

Here we test a parallelization scheme for shared memory architectures where we split the bounding box into smaller cubes for which the octree construction and integration can be performed separately from each other. The cubes are queued up and processed by a pool of worker threads. Finally the integration results from each worker thread are added up and the integral is converted to the Born radius via Eq. (2).

For assessing the accuracy of our proposed method we compare our computed Born radii using the solvent accessible surface[16] to those computed by solving the Poisson Boltzmann equation numerically using APBS[17] (see Fig. 2).The results from our method are in very good agreement with PB Born radii. The deviations are mainly attributed to the integral expression (2) itself, as was found in Ref. 6.

Timings and speedup measurements were run using the 70S Ribosome-tRNA Complex[18] with 220387 atoms on a node with four AMD Opteron 8431 six-core processors (see Fig. 3). We observe a very good scaling using up to 5 threads followed by a moderate speedup using up to 17 threads, reducing the computation time to 11% of the single

PowerBornRadii Accuracy

PowerBornRadii multi-threaded version
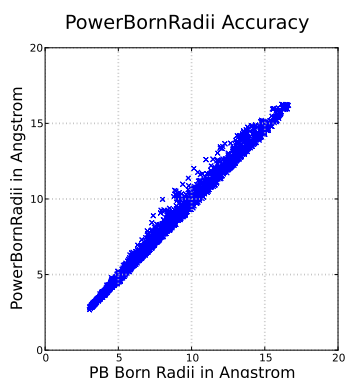
Figure 2. Precision of PowerBornRadii compared to those computed using the PB solver APBS[17] for the solvent accessible surface of PDB 1OCA.
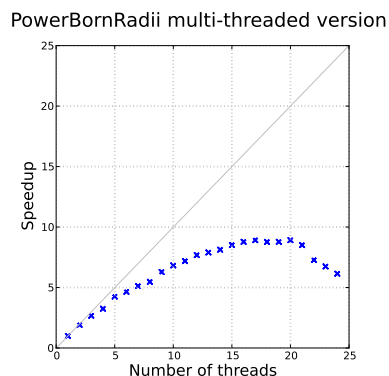
Figure 3. Speedup measurements for the multi-threaded version of PowerBornRadii on a 24 core computation node. For more details see the text.

threaded version and resulting in a total runtime of 0.62 seconds for the computation of 220387 Born radii. Using more threads leads to an increase in computation time, thus requiring more code optimization to further improve parallelization.

The present implementation will be made available, which will significantly speed up biomolecular simulations involving the Generalized Born model on multi-core shared memory machines, which are widely in use today. An MPI based implementation for distributed memory machines is under development.

Acknowledgments

References

1. Jason Wagoner and Nathan A Baker, *Solvation forces on biomolecular structures: A comparison of explicit solvent and PoissonBoltzmann models*, Journal of Computational Chemistry, **25**, no. 13, 1623–1629, Oct. 2004.
2. Jun Wang and Ray Luo, *Assessment of linear finite-difference Poisson-Boltzmann solvers*, Journal of Computational Chemistry, pp. NA–NA, 2010.
3. W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson, *Semi-analytical treatment of solvation for molecular mechanics and dynamics*, Journal of the American Chemical Society, **112**, no. 16, 6127–6129, 1990.
4. Alexey Onufriev, David A Case, and Donald Bashford, *Effective Born radii in the generalized Born approximation: the importance of being perfect*, Journal of Computational Chemistry, **23**, no. 14, 1297–1304, Nov. 2002, PMID: 12214312.

5. Tomasz Grycuk, *Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation*, The Journal of Chemical Physics, **119**, no. 9, 4817–4826, 2003.

6. John Mongan, W Andreas Svrcek-Seiler, and Alexey Onufriev, *Analysis of integral expressions for effective Born radii*, The Journal of Chemical Physics, **127**, no. 18, 185101, Nov. 2007, PMID: 18020664.

7. Di Qiu, Peter S. Shenkin, Frank P. Hollinger, and W. Clark Still, *The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii*, The Journal of Physical Chemistry A, **101**, no. 16, 3005–3014, Apr. 1997.

8. Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar, *Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium*, The Journal of Physical Chemistry, **100**, no. 51, 19824–19839, Jan. 1996.

9. Alexey Onufriev, Donald Bashford, and David A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model*, Proteins: Structure, Function, and Bioinformatics, **55**, no. 2, 383–394, Mar. 2004.

10. Harianto Tjong and Huan-Xiang Zhou, *GBr6: A Parameterization-Free, Accurate, Analytical Generalized Born Method*, The Journal of Physical Chemistry B, **111**, no. 11, 3055–3061, Mar. 2007.

11. Emilio Gallicchio, Kristina Paris, and Ronald M. Levy, *The AGBNP2 Implicit Solvation Model*, Journal of Chemical Theory and Computation, **5**, no. 9, 2544–2564, 2009.

12. Boris Aguilar, Richard Shadrach, and Alexey V. Onufriev, *Reducing the Secondary Structure Bias in the Generalized Born Model via R6 Effective Radii*, Journal of Chemical Theory and Computation, **6**, no. 12, 3613–3630, Dec. 2010.

13. Wei Cai, Zhenli Xu, and Andrij Baumketner, *A new FFT-based algorithm to compute Born radii in the generalized Born theory of biomolecule solvation*, Journal of Computational Physics, **227**, no. 24, 10162–10177, Dec. 2008.

14. Rezaul Chowdhury and Chandrajit Bajaj, *Multi-level grid algorithms for faster molecular energetics*, in: Proceedings of the 14th ACM Symposium on Solid and Physical Modeling, SPM '10, p. 147152, ACM, New York, NY, USA, 2010, ACM ID: 1839799.

15. Martin Brieg and Wolfgang Wenzel, in preparation, 2011.

16. B. Lee and F.M. Richards, *The interpretation of protein structures: Estimation of static accessibility*, Journal of Molecular Biology, **55**, no. 3, 379–400, IN3–IN4, Feb. 1971.

17. Nathan A. Baker, David Sept, Simpson Joseph, Michael J. Holst, and J. Andrew McCammon, *Electrostatics of nanosystems: Application to microtubules and the ribosome*, Proceedings of the National Academy of Sciences of the United States of America, **98**, no. 18, 10037 –10041, 2001.

18. Andrei Korostelev, Sergei Trakhanov, Martin Laurberg, and Harry F. Noller, *Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements*, Cell, **126**, no. 6, 1065–1077, Sept. 2006.

Biophysical and Computational Studies on the Bacterial Haem Chaperone, HemS

Desmond C. Y. Choy, Christopher S. Whittleston, Elizabeth B. Sawyer, David J. Wales, and Paul D. Barker

Department of Chemistry, University of Cambridge
Lensfield Road, Cambridge, CB2 1EW, United Kingdom
E-mail: {dc392, pdb30}@cam.ac.uk

The iron-porphyrin, haem, is a ubiquitous, versatile molecule essential to all life. Some organisms can synthesise haem but certain pathogenic bacteria sustain life using exogenous haem. HemS is a protein vital to *Yersinia enterocolitica* and contains the important missing link between our knowledge on haem acquisition and iron/haem utilization in Gram-negative bacteria. Our research combines biophysical and computational techniques to study the protein's function(s) and the mechanism(s) in achieving it/them, with recent experimental results suggesting that HemS is a novel NADH-dependent haem breakdown enzyme. The mechanistic details are studied computationally by the stochastic basin-hopping algorithm, employed here for the first time in a discovery-oriented application and has thus far revealed properties of the protein which have not yet been observed experimentally. This insight directs further experimental research and this dual approach has proven to be a powerful route to explore new mechanisms which were previously difficult to access.

1 Introduction

1.1 Iron and Haem in Biology

Iron is a versatile and abundant element. Often in biology, iron is found to be held in a versatile, asymmetrical and ubiquitous organic ring, the porphyrin – this is haem (*b*). This magnificent cofactor is responsible for many roles in biology, *e.g.* O_2 carrier in hemoglobin or electron conductor in certain cytochrome systems. It is difficult to define what a haem binding pocket is, because there are many ways to interact with this molecule. It is made even more difficult in the current research, because we are focusing on a class of proteins which are not the terminal receptors of haem as exemplified above, but a group of proteins which pass on haem from one place to another, a class known as the haem chaperones.

1.2 Haem Uptake System in *Yersinia enterocolitica*

Many organisms can synthesise haem by a well-studied biosynthetic pathway, or they can acquire haem from external sources. The pathogenic Gram-negative bacteria *Yersinia enterocolitica*, often found in raw meat and causes severe food poisoning, acquires haem by scavenging off the molecule from a haem-rich source, *e.g.* red blood cells. Once it acquires the cofactor, it passes haem onto a series of haem chaperones in order to permeate two membranes. This is the Hem uptake system[1]. The terminal receptor of this system is HemS, the focus of the current research. The *hemS* gene is essential to *Y. enterocolitica* and is a vital facilitator between the processes of haem uptake and iron release, but its precise role is currently unknown.

2   Biophysical Work on HemS

The crystal structure of HemS has been solved in both the *apo-* and *holo-*forms and it is shown to be unlike any other class of haem binding protein[2]. Despite the pseudo-C_2 symmetry, HemS is observed to bind only one haem in an unusual haem binding pocket and haem binding involves global conformational change.

One of the first tasks of this project was to elucidate the function of this protein. It was found that incubating *holo-*HemS in NADH yields a novel product, which is purple and not any product along the canonical haem oxygenase pathway[3]. There is no known system where NADH directly reacts with haem. Purifying and characterising this breakdown product is an ongoing effort. A puzzling feature of this reaction is that haem seems to inhibit its own breakdown, at a concentration where haem is known to exist primarily in the dimeric form[4]. Searching the surfaces of the HemS cavities against known NADH binding sites by Relibase[+5] reveals an NADH binding site adjacent to the haem binding pocket. Could haem, therefore, inhibits its own breakdown by blocking the NADH-binding site as a dimer? Even more interestingly, mass spectrometry data and mutagenesis studies have provided evidence that HemS can bind more than one haem molecule, but the second one binds with a much lower affinity.

3   How Computational Work Guides Experiments

Experimental data put the stoichiometry of HemS, as suggested by the crystal structure, into question. Furthermore, we need to address its NADH-dependent enzymatic activity. In other words, we want to elucidate the NADH binding mechanism, and to address the possibility of (an) alternative haem-binding site(s). This could be two sides of the same coin, or might not be related at all.

The experimental challenge is that reactive structures and low affinity binding sites are difficult to study physically. In order to study this problem at an atomic level, we approach it theoretically by examining the underlying potential energy surface (PES) of the system – we want to know how the energy of the system changes with respect to the nuclear geometry of the system. This is performed using the state-of-the-art software, GMIN (interfaced with AMBER9)[6], which stochastically explores and locates energy minima of the PES using the basin-hopping algorithm[7]. This is the first time this method has been applied to a major discovery-oriented application.

4   The NADH Binding Site

Bioinformatics has given us a starting point for calculations, and as such, the NADH conformation from Relibase[+] was superimposed onto the *holo-*HemS structure and was subjected to energy minimisation and geometry optimisation using (AMB)GMIN. In the software, the user can specify moves or combinations of moves to perturb the protein and/or ligand structures to sample the configuration space of the system. A "basin-hopping step" can include, therefore, moves such as rotating the side chains surrounding the ligands, rotating around any bonds of the ligand itself, overall translation and rotation of the ligand, or even performing molecular dynamics on the entire system. This overcomes the common
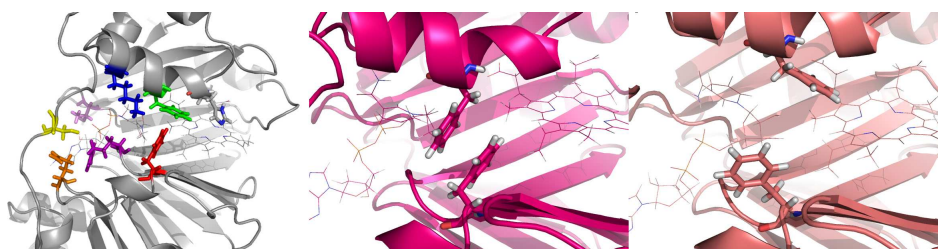
Figure 1. (*a*) Typical Result from *holo*-HemS+NADH Calculations; (*b*) and (*c*) Two examples of the conformations the Phe gate adopt.

problem of conventional bioinformatics methods where global conformational changes of protein upon ligand binding is often difficult to address.

HemS is a large system, and so these calculations are computationally very expensive, but from these we have thus far observed a number of interesting structural features of the protein which appear to play important roles in NADH binding and activity. Shown in Fig. 1(*a*) is a typical result of such calculations and seven residues are highlighted. These residues are frequently identified as key functional groups for NADH recognition, especially at the phosphate backbone of the ligand, and such structural motifs have never been seen before. In particular, the lysine (blue) is often responsible for amide recognition, which is coupled with, shown in Figures 1(*b*) and (*c*), an unusual pair of flexible phenylalanine residues which often forms a gate in different conformations shielding the haem from the rest of the cavity. BLAST search showed that these residues are well-conserved in homologous proteins. It is possible that π-π interactions play a key role in the NADH reaction, given the nicotin-amide becomes charged-aromatic upon hydride transfer.

Superimposing the NADH conformations of all calculations show good structural convergence overall, especially at the phosphate backbone, with the putative reactive functional group, the nicotin-amide, adopting multiple possible conformations. When the same calculations were repeated on the *apo*-structure, there is little structural convergence in protein and ligand, suggesting the necessity for the presence of haem to activate the cavity (see Fig. 2).
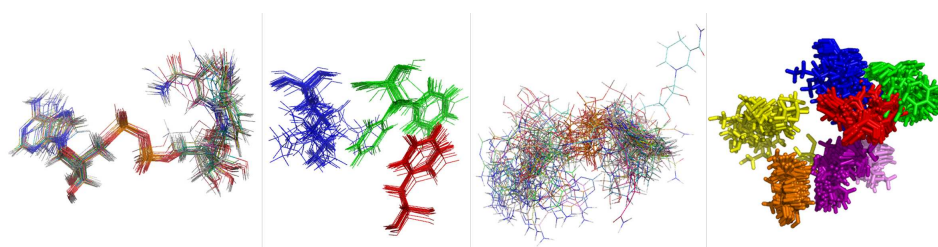
Figure 2. Results of parallel calculations: (*a*) NADH conformations of *holo*-HemS; (*b*) residue conformations of *holo*-HemS; (*c*) NADH conformations of *apo*-HemS; (*d*) residue conformations of *apo*-HemS.

5 Conclusions and Future Work

HemS is an NADH-dependent haem breakdown enzyme using unprecedented chemistry to release iron from haem in a haem concentration-dependent manner. The basin-hopping algorithm has enabled us to confidently identify an NADH binding site in HemS and suggest seven key residues that could be crucial for NADH recognition. Consequently we are currently in a position to validate this experimentally by mutagenesis. It is also important to identify the structure of the haem breakdown product. The issues of substrate inhibition and an alternative haem binding site are being addressed by the same method, and is currently work in progress.

We believe that HemS is a regulatory protein which serves a dual role in haem sequestration and haem breakdown / iron release depending on the physiological environment. The novel computational method enabled by GMIN has provided much valuable insights into this problem which would otherwise be impossible to identify experimentally alone. The combination of bioinformatics, theory and experiments has proven to be a powerful route to explore new mechanisms which were previously difficult to access.

References

1. I. Stojiljkovic, and K. Hantke, *Transport of haemin across the cytoplasmic membrane Haem Binding to HemS through a haemin-specific periplasmic binding-protein-dependent transport system in Yersinia enterocolitica*, Mol. Microbiol. **13**, 719, 1994.

2. S. Schneider, K. H. Sharp,, P. D. Barker, and M. Paoli, *An induced fit conformational change underlies the binding mechanism of haem-transport proteobacteria-protein HemS*, J. Biol. Chem. **281**, 32606, 2006.

3. M. Unno, T. Matsui, and M. Ikeda-Saito, *Structure and catalytic mechanism of heme oxygenase*, Nat. Prod. Rep. **24**, 553, 2007.

4. K. A. de Villiers, C. H. Kaschula, T. J. Egan, and H. M. Marques, *Speciation and structure of ferriprotoporphyrin IX in aqueous solution: spectroscopic and diffusion measurements demonstrate dimerization, but not -oxo dimer formation*, J. Biol. Inorg. Chem. **12**, 101, 2006.

5. Relibase$^+$ - Easy Searching of Protein-Ligand Complexes,
   `http://www.ccdc.cam.ac.uk/products/life_sciences/relibase`.

6. GMIN: A program for finding global minima and calculating thermodynamic properties from basin-sampling, `http://www-wales.ch.cam.ac.uk/GMIN/`.

7. D. J. Wales, and J. P. K. Doye, *Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms*, J. Phys. Chem. A **101**, 5111, 1997.

Molecular Dynamics Studies of the Human Transporter Associated with Antigen Processing (TAP)

Valentina Corradi and D. Peter Tieleman

Institute for Biocomplexity and Informatics, Dept. of Biological Sciences, University of Calgary
2500 University Dr NW, Calgary, AB, Canada T2N 1N4
E-mail: {vcorradi, tieleman}@ucalgary.ca

ATP-binding cassette (ABC) transporters are one of the largest and ubiquitous families of membrane proteins. ABC transporters translocate a variety of substrates across membranes by undergoing conformational changes induced by the binding and hydrolysis of ATP molecules. Even though different subfamilies of these transporters have been identified, eukaryotic and prokaryotic transporters share a common structural architecture, which consists of two transmembrane domains (TMDs), and two nucleotide binding domains (NBDs). The human transporter associated with antigen processing (TAP) is a heterodimeric ABC transporter that shuttles antigenic peptides from the cytosol into the endoplasmic reticulum lumen. We used molecular dynamics simulations to investigate the dynamics of this transporter as a function of the nucleotide-bound state. The ATP-bound system shows a high structural stability, while larger fluctuations characterized the ADP-bound state, resulting in a partial opening of the NBD dimer in some of the simulations.

1 Introduction

ABC transporters couple the hydrolysis of ATP with the transport of substrates across membranes. They consist of two nucleotide binding domains (NBDs), where ATP binding and hydrolysis occur; two transmembrane domains (TMDs), which constitute the translocation pathway, and optional additional domains. Two ATP binding sites are located at the interface between the NBDs, and consist of highly conserved motifs involved in ATP binding. Because of the head-to-tail orientation of the NBDs, each ATP binding site is formed by the Walker A, Walker B motif, as well as H- and Q-loop from one NBD combined with the Signature motif and the D-loop from the second NBD. The ATP binding induces NBD dimerization, and the following ATP hydrolysis provides the energy needed for the conformational changes of the TMDs[1].

TAP is an ABC transporter involved in the immune response by translocating antigenic cytosolic peptides into the endoplasmic reticulum for loading into MHC class I molecules[2]. TAP is a heterodimer (TAP1/TAP2), characterized by asymmetric ATP-binding sites (Fig. 1). As a consequence of the mutations occurring in the NBDs, the ATP binding site in TAP1-NBD (TAP1-Walker A and B motifs, TAP2-Signature motif) is degenerate, and shows a lower ATPase activity with respect to the consensus site in TAP2-NBD (TAP1-Signature motif, TAP2-Walker A and B motifs). In this study, we investigated the dynamics of TAP-NBDs by means of molecular dynamics (MD) simulations. We built a model for TAP-NBDs, and we simulated the system in the presence of ATP or ADP molecules.
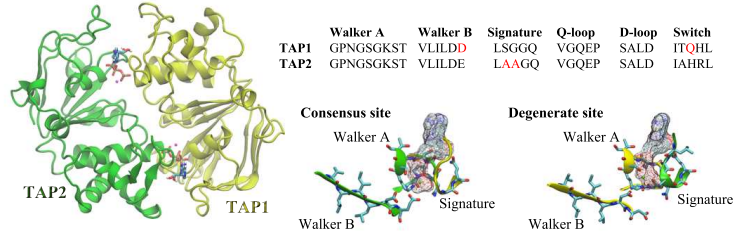
Figure 1. Ribbon representation of the TAP-NBD dimer. TAP1 and TAP2 are in yellow and green, respectively. The residues forming the conserved motifs in the TAP1- and TAP2-NBD are displayed, and the non-consensus residues are highlighted in red. A close up view of the consensus and the degenerate sites is also shown. The nucleotide is represented as mesh.

2 Methods

Homology modeling. A model of the TAP-NBDs was built by first performing a sequence alignment between these two domains using ClustalW[3]. The crystal structure of TAP1-NBD (PDB access code 1JJ7) was then used as a template to build a 3D TAP2-NBD model using with Modeller 9v6 program[4]. TAP1- and TAP2-NBD were then aligned to Sav1866-NBDs, to obtain the final TAP-NBD dimer. ATP, ADP and Mg^{2+} were added according to PDB codes 2HYD and 2IXE, respectively.

MD simulations. All simulations were carried out with GROMACS 4.0.7 software, using the GROMOS53A6 force field[5]. Each system (ATP1-ATP2, and ADP1-ADP2) was minimized first with position restraints for protein heavy atoms and nucleotide atoms, then with position restraints for the nucleotide. 1 ns equilibration was performed with position restraints for protein heavy atoms. Per each system, 4 independent simulations (R1, R2, R3, R4) of 100ns long were ran.

3 Results

To describe the dynamics of TAP-NBDs, we measured the distances between the β-phosphate of ATP or ADP and the center of mass (COM) of the Signature motif, as well as the distance between the COM of the Signature and the Walker A motif. We also calculated the root mean square fluctuations (RMSF) of the nucleotide (ATP or ADP) and the Signature and Walker A motifs in each binding site. The comparison of the values obtained from the ATP1-ATP2 system with the values obtained from the ADP1-ADP2 system describes possible structural rearrangements on the NBDs as a consequence of ATP hydrolysis.

Fig. 2A-B shows the results for the degenerate site. When ATP was present (Fig. 2A), no significant differences in the measured distances were observed. After 100 ns of simulation, the mutual orientation of the nucleotide and the surrounding motifs remained unaltered in all the four runs. In contrast, the presence of ADP (Fig. 2B) increased the distance between the β-phosphate of ADP and the COM of the Signature motif by about 0.2 nm during the first 20 ns. The same pattern was found for the distance between the COM of the
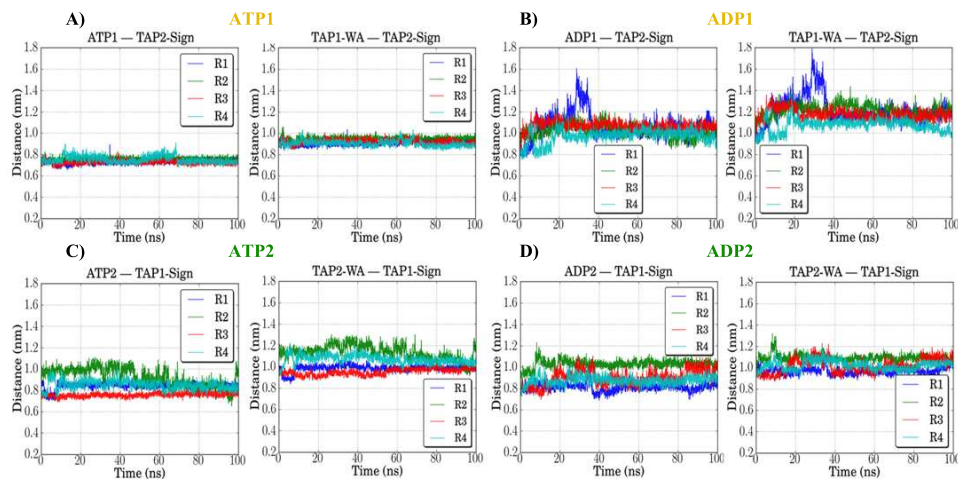
Figure 2. A-B. Degenerate site. The distances between the β-phosphate of the nucleotide (A, ATP, and B, ADP) as well as the distances between the Walker A of TAP1 and the Signature motif of TAP2 are shown as a function of the simulation time. C-D. The corresponding distances are shown for the consensus site in the presence of ATP (C) and ADP (D). Sign, Signature motif; WA, Walker A motif.

Signature and the Walker A motifs. In these simulations, ADP stayed bound to the Walker A motif, while the Signature motif adopted a different orientation with respect to the starting structure. The RMSFs for the degenerate site when ADP was present were higher than the values found for the ATP1-ATP2 system (Fig. 3). Fig. 2C-D displays the results for the consensus site. In the presence of ATP (Fig. 2C), for three runs (R1, R2, and R4) the measured distances were higher than the values obtained from ATP1 in the degenerate site (Fig. 2A). However, after 60 ns, the distances converged to values close to the initial ones, and no dissociation between TAP1 and TAP2 was observed. In contrast, the presence of ADP destabilized the dimer interface. The distance between the ADP β-phosphate and the Signature motif increased of 0.05 (R1), 0.1 (R4) or 0.2 (R2, R3) nm after 100 ns, indicating a closer interaction of ADP with the Walker A motif (Fig. 2D, left). Furthermore, after 100 ns, the distance between the COM of the Signature motif and the Walker A motif is 0.1 nm higher than the starting structure (Fig. 2D, right). As for the degenerate site, the RMSFs of the nucleotides and the conserved motifs were higher when ADP was present (Fig. 3).

4   Conclusions

Molecular dynamics simulations were used to investigate the dynamics of the TAP-NBD heterodimer in the presence of either ATP or ADP molecules. For the ATP1-ATP2 system, the interactions between ATP, the Signature and the Walker A motif are preserved both in the consensus and the degenerate sites for the whole length of the simulations. No dimer opening is observed in this set of simulations. However, the consensus site shows larger fluctuations and changes in the measured distances for the first 60 ns, perhaps due to the homology modeling of TAP2. The ADP1-ADP2 system is associated with a higher structural instability, and in this case, the degenerate and the consensus site show a different
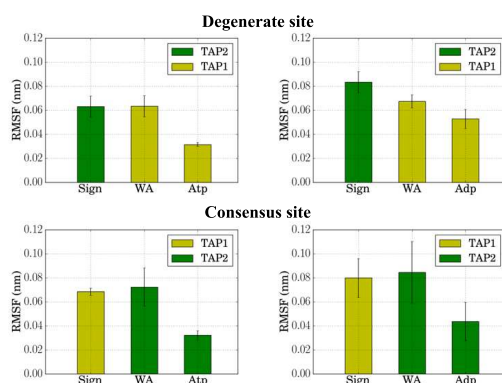
Figure 3. Root mean square fluctuations (RMSF) calculated for the nucleotides in the degenerate (upper graphs) and the consensus (lower graphs) site. These values are the average values obtained from the four independent runs per each system. Sign, Signature motif; WA, Walker A motif.

behavior. A clear increase in the distance between the Signature and the Walker A motif is observed only for the degenerate site. However, the lack of the γ- phosphate induces a different orientation of the nucleotides in both the binding sites with respect to the starting structure. ADP molecules stay bound to the Walker A motif, while most of the interactions with the Signature motif are not preserved. Additional simulations are in progress to confirm the partial opening of the NBDs in the presence of ADP.

Acknowledgments

References

1. E. Procko, M. L. O'Mara, W. F. Drew Bennett, D. Peter Tieleman, and R. Gaudet, *The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter*, FASEB J. **23**, 1287–1302, 2009.

2. M. Nijenhuis, and G. J. Hammerling, *Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site*, J. Immunol. **157**, 5467–5477, 1996.

3. J. D. Thompson,D. G.Higgins, and T. J. Gibson, *Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matric choice*, Nucleic Acids Res **22**, 4673–4680, 1994.

4. A. Sali, and T. L. Blundell, *Comparative Protein Modelling by Satisfaction of Spatial Restraints*, J. Mol. Biol. **234**, 779–815, 1993.

5. B. Hess, C. Kutzner, Carsten, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput. **4**, 435–447, 2008.

RNA Folding Dynamics Studied with Structure-based Models

Michael Faber and Stefan Klumpp

Max Planck Institute of Colloids and Interfaces, Science Park Golm, 14424 Potsdam, Germany
E-mail: michael.faber@mpikg.mpg.de

RNA molecules form three-dimensional structures as complementary bases form bonds and the molecule coils. These structures determine the function and biochemical activity of the molecule. For example, the presence or absence of a specific RNA structure can invoke transcriptional pauses or terminate the transcription altogether. We have developed a structure-based model for studying the folding dynamics of RNA secondary structures. To simulate the dynamics, we use a Monte-Carlo method with Metropolis rates, where the basic steps are the closing or opening of one native contact. We apply this model to the folding and unfolding of simple RNA structures in the presence and absence of an external force.

1 Introduction

RNA is a linear polymer made out of four different bases: adenine (A), cytosine (C), guanine (G) and uracil (U). Two pieces of an RNA molecule can connect via hydrogen bonds between complementary bases (AU and GC), such that the RNA folds into a three dimensional structure. RNA structures are usually described as a hierarchy of structures: the sequence of bases in the molecule is called the primary structure, the set of all base pairings the secondary structure and the three-dimensional shape of the molecule including all other structural elements the tertiary structure. Typically, this hierarchical description reflects the hierarchy of the folding process where the primary structure determines the secondary which in turn determines the tertiary contacts[1]. Great effort has been done on understanding and predicting the secondary structure of RNA molecules[2]. Moreover, over the last decade single molecule experiments using optical tweezers were performed on a number of RNA structures to study their stability, their force dependence and their dynamics. In cells, RNA structures often fold while the RNA is transcribed. In such cases, the dynamics of folding is typically crucial for the function of the RNA. An example is the formation of hairpins during transcription, which can invoke transcriptional pauses or terminate the transcription altogether[3].

In the following we present a simple model which aims at describing the dynamics of RNA secondary structures. The model we have developed is a structure based model, i.e. we concentrate on the native contacts of a given RNA structure and study its dynamics. Structure-based models have been used extensively in studies of protein folding. They are based on the principle of minimal frustration that states that functional sequences have been selected to avoid energetic frustration to ensure rapid folding[4]. As a consequence, the dynamics of folding is expected to be governed by the same interactions that govern the folded state. The same arguments should also apply to the folding of structured RNAs, and indeed similar argument have occasionally been used for RNA[5]. Here we use a Monte Carlo method to simulate the folding dynamics. We will show results on the stability of

secondary structures and the distributions of folding and unfolding times as well as on force induced unfolding.

2 Model

The secondary structures we consider consists of five basic structural motifs that arise from base pairing: simple unconfined single stranded pieces of RNA, helical regions of subsequent paired bases, hairpin loops that form an end to a helical region, internal loops with more then one outgoing helical region and bulges. If we number the bases of an RNA molecule $\{1 \dots N\}$ the secondary structure can be described as a set of pairs (i, j) denoting the formed base pairs. Here we consider only structures without pseudo knots which is a common restriction in the prediction of secondary structures. Therefore two base pairs (i, j) and (i', j') must either fulfill $i < i' < j' < j$ or $i' < i < j < j'$. These conditions ensure that no base pair can form between a base in the region separated by the first base pair (i, j) and a base outside that region.

Our RNA model is structure based. We take the RNA to be a sequence of bases where only specific, predefined contacts can be made between bases of that sequence. These positions are defined by the native (folded) structure of the RNA molecule. Here we restrict ourselves to contacts that form by base pairing, but additional types of contacts could also be included. Then the dynamics of the RNA molecule are analyzed using a Monte-Carlo method with Metropolis rates. The basic steps are the closing and opening of contacts. This is done by choosing a base pair randomly from the list of possible base pairs. If the chosen base pair exists already, then it might open, and if it does not exist, it may close. The probabilities for the opening or closing moves are calculated from the free energy difference of the structure before and after the step. We assume that the free energy of a structure can be calculated as a sum of energy contributions from the different structural motifs. Forming a base-pair is energetically favorable, on the other hand the formation of a loop constrains the RNA molecule which is entropically costly

$$G_{\text{tot}} = \sum_{\text{all basepairs}} G_{\text{basepair}} + \sum_{\text{all loops}} G_{\text{loop}}. \tag{1}$$

Here we use a simple parametrization of the free energies: We chose each base pair to contribute $G_{\text{basepair}} = -2$ kcal/mol. Energy contributions of loops depend logarithmically on the loop length. For hairpin loops, which require $n \geq 3$ bases in the loop, we take $G_{\text{hairpin loop}}(n) = \left(5 + \ln(n/3)\right)$ kcal/mol, while internal loops and bulges are assigned $G_{\text{int loop}}(n) = \left(2 + ln(n)\right)$ kcal/mol.

3 Simulation Results

In the following we will use a contrived and simple hairpin structure to demonstrate key features that our model describes. Our model structure consists of a loop closed by consecutive identical base pairs. Despite its simplicity, this structure already shows some universal properties which one can expect to find in more complicated systems. First we look at the behavior of a free hairpin (Fig. 1). Starting simulations with a fully closed structure, i.e. with all possible base pairs formed, we observe base pairs to open, and after some time
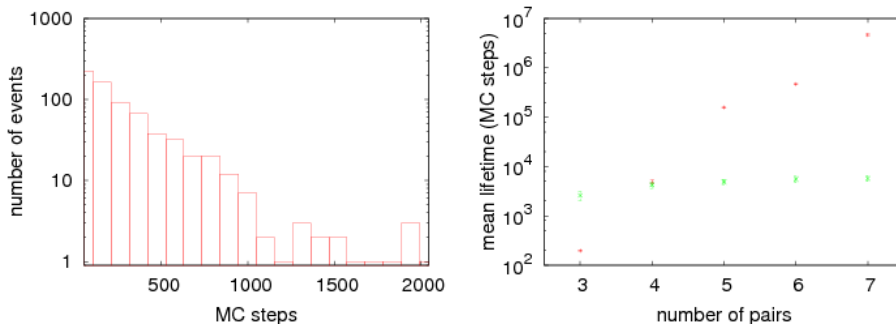
Figure 1. Lifetime (unfolding time) of a hairpin: left: Distribution of the lifetime of a hairpin with a 5 bp stem-The lifetime is defined as the time it takes to get from the state where all possible base pairs are closed to the state where all are open. right: Mean lifetime plotted as a function of the number of base pairs in the stem (red) and corresponding closing (folding) time (green).

all contacts are disconnected for the first time. We call this the lifetime or unfolding time of the hairpin. The distribution of unfolding times is exponential (Fig. 1(left)), a hallmark of two-state folding, as also indicated by experiments[1]. Likewise the distribution of folding times is also exponential. We then varied the length of the stem, i.e. the number of possible base pairs in the stem. The mean lifetime of the hairpin depends exponentially on the stem length, while the folding time, the time it takes to close a hairpin from a single stranded chain does not depend strongly on the stem length (Fig. 1(right)). This result is plausible since the limiting step of folding is the formation of the first bond, which is unfavorable due to the loss of entropy from the loop formation, while the other base pairs are closed very quickly once the first bond is formed.

Next, we use our model to simulate a hairpin under pulling forces. We introduce an additional energy term which goes with $F_{ext} * \Delta x$, where F_{ext} is a constant external force and Δx the relevant change in chain length arising from base pairing. We determine the equilibrium distribution between the folded and unfolded state as a function of the applied force. For a hairpin of length 5, we observe a sharp transition from mostly closed to mostly open at about 9 pN (Fig. 2), reminiscent of experimental observations for more complex hairpins[6].

4   Concluding Remarks

We have studied the folding and unfolding dynamics of simple RNA molecules with Monte Carlo simulations of a structure based model. With our model we are able to show the expected dynamic behavior of an RNA hairpin. As may be expected we find folding and unfolding times that are exponentially distributed. The folding of such a structure is mainly limited by the formation of the first base pair, while the dissolution strongly depends on the length of the stem which gives the stability of the folded state. Our model also allows us to introduce external forces on the RNA molecule. We see a typical force extension behavior where at a narrow force range the RNA changes from a folded to an unfolded formation. This model can also be extended towards a more detailed and more realistic, empirical
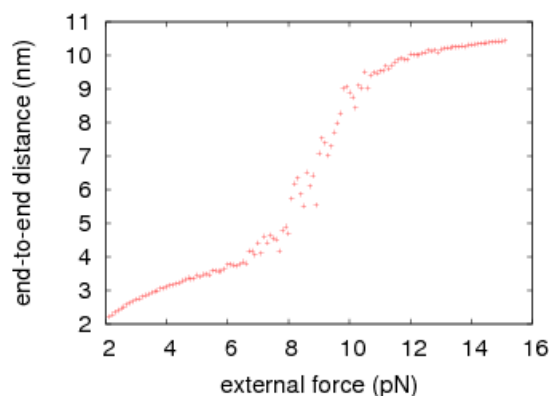
Figure 2. Mean end-to-end distance of an RNA molecule as a function of the external pulling force. Our model hairpin of length 5 bp unzips.

energy parametrization, similar to what is used in secondary structure prediction. With that parametrization, which is, of course, sequence dependent, quantitative agreement with the experimental data is obtained[7].

References

1. I Tinoco and C Bustamante, *How RNA folds.*, J Mol Biol, **293**, no. 2, 271–81, Oct. 1999.
2. M Zuker, *Mfold web server for nucleic acid folding and hybridization prediction.*, Nucleic Acids Res, **31**, no. 13, 3406–15, July 2003.
3. T Henkin and C Yanofsky, *Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions.*, Bioessays, **24**, no. 8, 700–7, Aug. 2002.
4. 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.
5. D Pincus, S Cho, C Hyeon, and D Thirumalai, *Minimal models for proteins and RNA from folding to function.*, Prog Mol Biol Transl Sci, **84**, 203–50, 2008.
6. J Liphardt, B Onoa, S B Smith, I Tinoco, and C Bustamante, *Reversible unfolding of single RNA molecules by mechanical force.*, Science, **292**, no. 5517, 733–7, Apr. 2001.
7. M Faber and S Klumpp, *Manuscript in preparation*, nn, 2011.

Modeling Protein Structures and their Complexes with Sparse Experimental Data

Dominik Gront, Maciej Błaszczyk, Jacek Wabik, and Andrzej Kolinski

University of Warsaw, Faculty of Chemistry, Pasteura 1, 02-093 Warsaw, Poland
E-mail: {dgront, mblaszczyk, jwabik, kolinski}@chem.uw.edu.pl

BioShell project has been started in 2005 as a set of stand-alone programs aimed on simplification of typical bioinformatics tasks. Since then it has evolved to become a fully featured scripting language for biomolecular modeling and structural bioinformatics. Most recently, the development of the package is focused on incorporating various types of experimental data into protocols for structure prediction of proteins and their complexes. In this work we present an application of Small Angle Xray Scattering (SAXS) profiles to the determination of mutual domain orientation in multi-domain proteins. Preliminary results suggest that the scattering data can be successfully used in studies of large macromolecular assemblies providing that the structure of the individual interacting partners are known.

1   Introduction

One of the most important challenges in modern structural biology is the characterization of multi-domain macromolecular complexes that govern a major part of important cellular functions. These large biomachines are very difficult targets for standard experimental methods. Due to their size and flexibility investigation with X-ray crystallography or NMR spectroscopy is not always easily accomplishable. Multidisciplinary methods are amongst the most promising approaches. Structures of separate protein domains, that have been previously determined with Xray or NMR methods may be properly combined into whole complexes with the help of Electron Microscopy (EM) or Small Angle Xray Scattering (SAXS) data. SAXS profile is a function of all atomic coordinates for a given system and provides only an averaged description of macromolecular size and shape encoded in a very synthetic way. Informational entropy analysis suggests that a scattering profile may be used to determine only several independent degrees of freedom. The data however is considered to be satisfactory for the unique definition of geometry of a macromolecular assembly.

SAXS data has been already incorporated into many modeling platforms, e.g. NIH-Xplor[1] or ATSAS[2]. Here we describe its successful combination with the BioShell modeling platform. BioShell[3] package has been originally created for structural bioinformatics[4]. The suite of programs was growing and new functionalities have been emerging. After several years of development the package is capable to deliver all the necessary modeling routines[5]. BioShell routines may be called from any programming language that operates within Java Virtual Machine (JVM), most notably from Java implementations of Python (jython) and Rubby (jRuby) as well as from Java itself. This makes it a very versatile platform that may be quickly applied to various modeling projects.

2 Materials & Methods

In this contribution SAXS data has been used in determination of three-dimensional (3D) structure of two-domain proteins, provided that the high-resolution structures of both domains are known. Conformation of the linker that connects domains has been subjected to conformational sampling while the structure of each of the domains has not been altered. SAXS data has been applied as the only scoring term to guide the walk in the conformational space towards the correct geometry.

2.1 Conformational Sampling

CartesianProteinSystem module of the BioShell package has been used for conformational sampling. Due to its very general and careful design, a wide range of mover objects are available for introducing conformational changes. In the course of this work, RandomDihedralMover has been used to modify Φ,Ψ dihedral angles at randomly selected position in the linker region. The mover object proposes new Φ,Ψ coordinates according to an empirical probability distribution that has been obtained from loop conformations extracted from a non-redundant set of high-resolution protein structures. A trial conformation has been accepted or rejected according to the Metropolis criterion with SAXS-based χ^2 statistics. Simulated annealing Monte Carlo protocol was used to generate short trajectories, starting from a random conformation of the linker. The structure of the lowest energy (i.e. of the best χ^2 fit) has been reported from each trajectory.

2.2 SAXS-based Model Assessment

In this work we followed the frequently used approach to simulate SAXS intensity $I(q)$ at any arbitrary scattering vector length q with the Debye formula:

$$I(q) = \sum_{i=1}^{N_{at}} \sum_{j=1}^{N_{at}} f_i(q) f_j(q) \frac{sin(qd_{ij})}{qd_{ij}} \tag{1}$$

where d_{ij} is the distance between i-th and j-th atom of the molecule. Atomic form factors $f(q)$ were properly corrected to reflect the effect of displaced solvent. Following the work by Fraser et al.[6], dummy solvent atoms were placed at all atomic positions within the macromolecule with the form factors computed from on average electronic density of bulk water. The Debye formula was employed in computations of a theoretical SAXS spectrum for each trial protein conformation, i.e. at every Monte Carlo move. The quality of the fit between the experimental scattering data and those predicted from the models is described by the χ^2 statistics over the set of N_q values:

$$\chi^2 = \frac{1}{N_q - 1} \sum_{k=1}^{N_q} \left[\frac{I_{\text{reference}}(q) - I_{\text{model}}(q)}{\sigma(q)} \right] \tag{2}$$

| | | linker residues | |
PDB id	N_res	from	to
1a62A	125	42	47
1d09B	153	94	100
1mgtA	169	49	61
1ammA	174	79	88
1nkrA	195	96	102
1knyA	253	123	127
1ctuA	294	170	188
1j8mF	295	83	95
1jpnA	296	86	98
1calA	370	244	256
1bagA	425	344	349
1eovA	487	128	137

Table 1. Benchmark set used in this study. Each of these proteins comprises two domains. Conformational sampling has been restricted to the inter-domain linker region defined in the table.

3   Results

A benchmark set of twelve proteins (summarized in Tab. 1) have been utilized to test the protocol. These targets have been chosen to cover the typical range of polypeptide chain lengths, at the same time providing representatives for all the major protein architectures. For each of these proteins 50 000 structural models have been calculated. Results have been summarized in the Fig. 1 where each dot represents a single model. The plotted range of the coordinate root mean square deviation (crmsd) between a model and the native
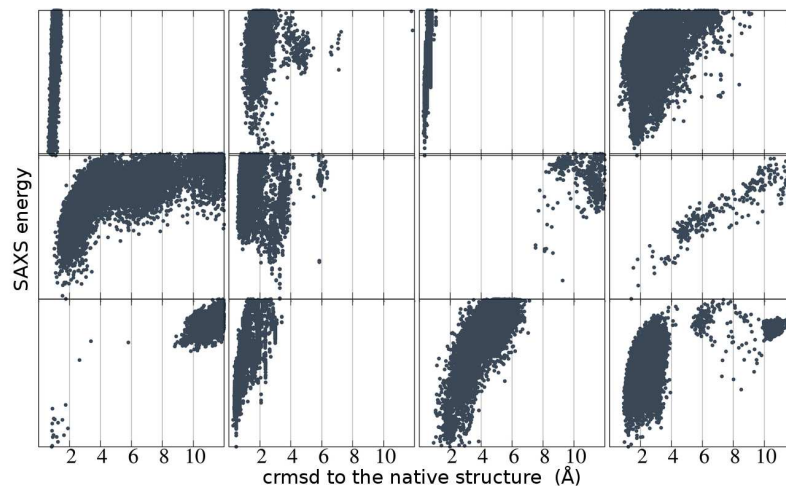
Figure 1. Scatterplots that show modeling results for the twelve proteins in the benchmark set. SAXS energy (Y axis, arbitrary units) is plotted as a function of crsmd (from 0.0 to 12.0 Å in each of the scatter plots).

structure (X axis) has been set to [0.0 Å,12.0 Å] although for some of the cases much larger values have been observed. Since SAXS scattering intensity $I(q)$ grows quadratically with the number of atoms in the scattering system, χ^2 statistics also varies greatly with the size of a target protein. Therefore the SAXS energy values on the Y axis are shown in arbitrary units, scaled separately for each box of the multipart plot. In all but one of the test cases (`1d09B`), SAXS energy decreases as the conformational sampling is approaching the native conformation according to a funnel-like dependence. Additionally, for two other cases (`1eovA` and `1j8mF`) conformational sampling turned out to be inefficient in the proximity of the native structure. Domain packing in the two other test cases: `1a62A` and `1mgtA` is relatively tight which explains the fact that the sampling process yielded mostly very good structures. Some conformations that otherwise could have low SAXS energy were excluded due to steric clashes.

4 Conclusions

From the perspective of typical biomolecular modeling methods, the test systems presented in this contribution should be considered as quite large. The problem has been made computationally tractable by freezing all but a small fraction of the degrees of freedom. Only from 12 to 36 main chain dihedral angles have been subjected to conformational sampling. This however granted enough flexibility to the protein chain to sample the whole space of mutual orientations between the domains. In general, SAXS - based score has been able to pinpoint the correct conformation. It should be also mentioned that even very small deviation in domain orientation (e.g. a tiny rotation of one of the domains in respect to the other one) may result in relatively large crmsd value.

Acknowledgments

References

1. A. Grishaev, and J. Wu, and J. Trewhella, and A. Bax *Refinement of Multidomain Protein Structures by Combination of Solution Small-Angle X-ray Scattering and NMR Data* J. Am. Chem. Soc **127**, 16621-16628, 2005.
2. M. Petoukhov, and P. Konarev, A. Kikhney, and D. Svergun *ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis.* J. Appl. Cryst. **40**, 223-228, 2007.
3. `http://www.bioshell.pl`
4. D. Gront, and A. Kolinski *BioShell - a package of tools for structural biology computations*, Bioinformatics **22**, 621-622, 2006.
5. D. Gront, and A. Kolinski *Utility library for structural bioinformatics*, Bioinformatics **24**, 584-585, 2008.
6. R. Fraser, and T. Macrae, and E. Suzuki *TITLE GOES HERE* J. Appl. Crystallogr **11**, 693-694, 1978.

HAMP Domain Region of Sensory Rhodopsin Transducers

Ivan Yu. Gushchin[1,2], **Valentin I. Gordeliy**[1,2,3]**, and Sergei Grudinin**[4,5]

[1] LPM, Institut de Biologie Structurale J.-P. Ebel, CEA-CNRS-UJF, 38027 Grenoble, France

[2] Research-Educational Centre "Bionanophysics", MIPT, 141700 Dolgoprudniy, Russia

[3] Institute of Structural Biology and Biophysics (ISB-2), Research Centre Jülich
52425 Jülich, Germany

[4] NANO-D, INRIA Grenoble Research Center, 38334 St. Ismier, France

[5] CNRS, Laboratoire Jean Kuntzmann, BP 53, Grenoble, France
E-mail: sergei.grudinin@inria.fr, ivan.gushchin@ibs.fr

Archaeal transducers of the sensory rhodopsins phototactic signal possess two HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) domains – a notable distinction from bacterial chemoreceptors. Here, we analyze the HAMP domain region (including the inter-HAMP region enclosed between the HAMP domains) of *Np*HtrII and discuss its possible roles for signal transduction. We propose that the inter-HAMP region structure is a coiled coil. Molecular dynamics simulations show that the protomers prefer to be longitudinally shifted by about 1.3 Å, with the free energy penalty for the symmetric structure in range 1.2-1.5 kcal/mol, which means that the inter-HAMP region is asymmetric and bistable. Both flanking HAMP domains are mechanistically coupled to the inter-HAMP region, and are also asymmetric. The longitudinal shift in the inter-HAMP region results in the displacement of the cytoplasmic part by 8.6 Å relative to the transmembrane part in the membrane plane. The established properties suggest that (1) the signal may be transduced through the inter-HAMP domain switching and (2) the inter-HAMP region may exist to enable dimeric cytoplasmic parts of the transducers to come close to each other to form higher-order oligomers.

1 Introduction

Archaea employ sensory rhodopsins (SR) I and II as photoreceptors to move towards red light, which enables bacteriorhodopsin and halorhodopsin activity, and to avoid harmful near-UV light[1]. The phototactic signal is generated in SR and then relayed to the transducer protein (Htr), whose cytoplasmic tip forms a complex with CheA kinase and modulates its activity, similar to bacterial chemoreceptors[2]. Meanwhile the primary events in the signal transduction chain are thought to be understood[1,3], further propagation of the signal through the HAMP domain(s) is much less clear[4]. Here, we describe the analysis of the *Natronomonas pharaonis* HtrII (*Np*HtrII) HAMP domain region structure and properties, and propose its role in the signal transfer.

2 Methods

The thorough description of the employed methods may be found elsewhere[5]. Homology models of the HAMP1 and HAMP2 domains were built using MODELLER[6]. Initial models for the inter-HAMP region were built manually with the help of SAMSON software[7]. All molecular dynamics simulations were conducted using NAMD2 program[8] with
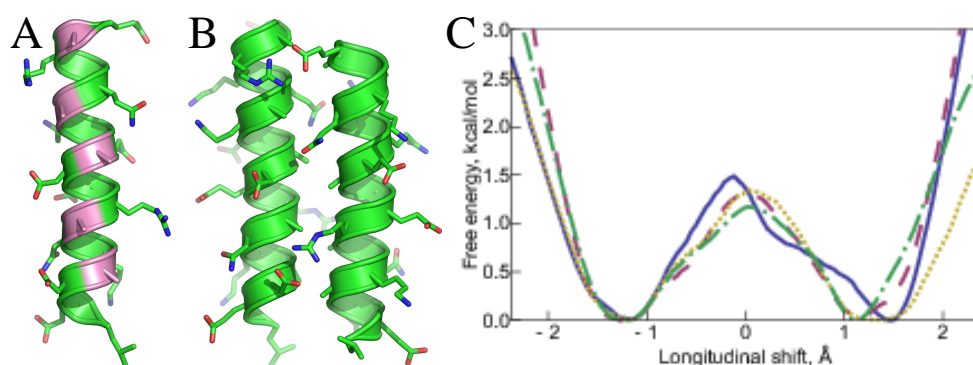
Figure 1. A) Hydrophobic groove of the *Np*HtrII inter-HAMP 135-153 fragment. Alanines are highlighted in pink. B) Model of a homodimer of the two *Np*HtrII inter-HAMP 135-153 regions. Note the asymmetric position of protomers and the differences in ionic bonds formed on each side of the structure. C) Free energy profiles for the relative longitudinal shift of the protomers in the inter-HAMP region (*Np*HtrII residues 135-153) at different molarities: 0.5 M (blue, solid), 1 M (magenta, dashed), 2 M (yellow, dotted) and 4 M (green, dot-dashed). Note that the minima are at about ± 1.3 Å and the barrier height is 1.2-1.5 kcal/mol. Reprinted with permission from Gushchin et al., Biochemistry 50, 574-580, 2011. Copyright © 2011 American Chemical Society.

CHARMM27 parameters. Preparations of initial models and data analysis using our own tcl-scripts were performed in VMD[9]. All residues were assumed to be in standard protonation conditions. Longitudinal shift was calculated as a difference between projections of the centers of mass of each helix on the common axis. To determine the free energy of a system as a function of longitudinal shift the umbrella sampling technique was used.

3   Inter-HAMP Region

Phototactic signal transducers have a notable distinction compared to bacterial chemoreceptors, as they have two HAMP domains instead of one. Their HAMP domains are separated by a short linker, with a typical length of two or four heptads. The structure of *Np*HtrII proteolytic fragment comprising residues 100-159 was determined by NMR spectroscopy to be α-helical[10]. Secondary structure predictions suggest that the linker forms a continuous α-helix with the second α-helix AS2 of HAMP1 and the first helix α-AS1 of HAMP2. Moreover, the linker is predicted to be in a coiled coil. Visual examination of the linker, which we call the inter-HAMP region, reveals that there is a hydrophobic groove on its surface, formed by alanines (Fig. 1A). Taking this all in consideration, it is highly reasonable to propose that the inter-HAMP region is a homodimeric coiled coil (Fig. 1B). Molecular dynamics simulations show that the monomeric inter-HAMP from *Np*HtrII, as well as the dimeric one, are stable. There is a notable feature in the dimeric inter-HAMP region, as there are longitudinal motions of protomers in the dimer. We analyzed this effect more thoroughly, determining the potential of mean force (PMF) for longitudinal displacements of protomers (Fig. 1C). The PMF shows that the system prefers the unsymmetrical conformation with the longitudinal shift of 1.3 Å (of either positive or negative sign). Thus, the inter-HAMP region is bistable.

Figure 2. A) Computed probability of finding the projection of the *Np*HtrII second HAMP domain center of mass on the membrane plane at the corresponding position, laid over the structure of *Np*SRI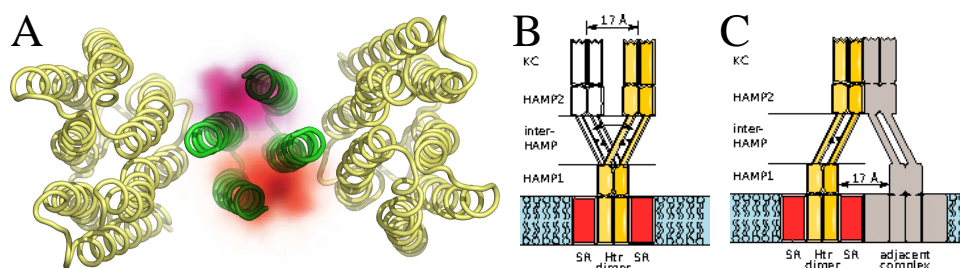I-*Np*HtrII 2:2 complex in a proper scale. View is perpendicular to the membrane plane. The initial structure was symmetric, and the tilt and the longitudinal shift in the inter-HAMP region have developed later. Two regions with the highest probability correspond to the two possible signs of longitudinal shifts and are colored red and magenta, correspondingly, to guide the eye. B, C) Two possible roles of the HAMP domain region of phototactic signal transducers. Model of the sensory rhodopsin transducer 2:2 complex is in yellow, with the protomers shown in different shades. Equivalent positions are marked on the inter-HAMP regions of the two protomers as black triangles to facilitate the perception of the longitudinal shift. B) Change of the longitudinal shift sign results in a distinct conformation, in which the kinase control module (KC) is displaced on average by 8.6 Å $\times 2 \approx 17$ Å along the membrane plane (second conformation is in white). Thus, the signal may be transduced through the inter-HAMP switching. C) Asymmetry of the HAMP domain region may facilitate the contacts with the adjacent transducers or receptors. Otherwise, these contacts would be impaired by bulky sensory rhodopsins residing in the membrane. Reprinted with permission from Gushchin et al., Biochemistry 50, 574-580, 2011. Copyright © 2011 American Chemical Society.

4 Model Including HAMP Domains

To determine whether the inter-HAMP region will keep its properties in presence of the flanking HAMP1 and HAMP2 domains, we expanded the model to include both of them. Initially prepared symmetric structure develops asymmetry during the simulation, with a longitudinal shift arising in the inter-HAMP region. Thus, the properties of the inter-HAMP region are conserved in physiological environment. Moreover, the inter-HAMP region affects HAMP domains, rendering them asymmetrical, as opposed to largely symmetrical models observed for sole HAMP domains[4]. The longitudinal shift in the inter-HAMP region seems to not affect the orientations of the HAMP domains, only displacing one relative to the other in the membrane plane. Analysis of the displacement reveals a shift of 8.6 ± 3.5 Å (Fig. 2A), which is a considerable value for the system.

5 Discussion

Signal transduction through HAMP domains is still not very well understood[4]. Here, we present the evidence, that in phototactic signal transducers, the region adjacent to HAMP domains (the inter-HAMP region) possesses interesting properties and is able to affect the HAMP domain structure. Asymmetry generated by the inter-HAMP region leads to displacement of HAMP2 in membrane plane relative to HAMP1. This raises an interesting possibility of signaling through the inter-HAMP switching (Fig. 2B). Such switching would generate a very large displacement of the cytoplasmic part. Another possible role of the HAMP domain region is to serve as a rigid mechanical joint that allows at the same time

the cytoplasmic parts to come close enough to each other to form higher-order oligomers (Fig. 2C). Such oligomers are generally considered to be necessary for proper signaling[2].

Acknowledgments

References

1. J. Sasaki and J. L. Spudich, *Signal Transfer in Haloarchaeal Sensory Rhodopsin-Transducer Complexes*, Photochem. Photobiol. **84**, 863-868, 2008.
2. G. L. Hazelbauer and W. C. Lai, *Bacterial chemoreceptors: providing enhanced features to two-component signaling*, Curr. Opin. Microbiol. **13**, 124-132, 2010.
3. R. Moukhametzianov, J. P. Klare, R. Efremov, C. Baeken, A. Goppner, J. Labahn, M. Engelhard, G. Büldt, and V. I. Gordeliy, *Development of the signal in sensory rhodopsin and its transfer to the cognate transducer*, Nature **440**, 115-119, 2006.
4. J. S. Parkinson, *Signaling Mechanisms of HAMP Domains in Chemoreceptors and Sensor Kinases*, Annu. Rev. Microbiol. **64**, 101-122, 2010.
5. I. Y. Gushchin, V. I. Gordeliy, and S. Grudinin, *Role of the HAMP Domain Region of Sensory Rhodopsin Transducers in Signal Transduction*, Biochemistry **50**, 574-580, 2011.
6. A. Sali and T. L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints*, J. Mol. Biol. **234**, 779-815, 1993.
7. S. Grudinin and S. Redon, *Practical modeling of molecular systems with symmetries*, J. Comput. Chem. **31**, 1799-1814, 2010.
8. J. C. 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.
9. W. Humphrey, A. Dalke, and K. Schulten, *VMD Visual molecular dynamics*, J. Mol. Graphics. **14**, 33-38, 1996.
10. K. Hayashi, Y. Sudo, J. Jee, M. Mishima, H. Hara, N. Kamo, and C. Kojima, *Structural Analysis of the Phototactic Transducer Protein HtrII Linker Region from Natronomonas pharaonis*, Biochemistry **46**, 14380-14390, 2007.

Energetics of Cross-reactivity of Diels-Alderase Antibody 1E9 and its Variants with Steroids

Parimal Kar, Reinhard Lipowsky, and Volker Knecht

Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany
E-mail: {Parimal.Kar, Reinhard.Lipowsky, Volker.Knecht}@mpikg.mpg.de

The energetics of binding of the steroid progesterone (PRG) and 5β-androstane-3,17-dione (5AD) to the Diels-Alderase antibody 1E9 or the $\text{Leu}^{\text{H47}}\text{Trp}/\text{Arg}^{\text{H100}}\text{Trp}$ 1E9 double mutant (1E9dm) has been investigated using the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method. The results have been compared to the energetics for the complex of PRG with the anti-progesterone antibody DB3. The steroid binds less strongly to 1E9 than to DB3 but the mutation improves the steroid affinity, in agreement with experiment. Although the complexes formed by PRG with 1E9dm and DB3 have similar affinity, the binding mechanisms differ. Reduced van der Waals interactions for PRG-1E9dm versus PRG-DB3 are energetically compensated by an increased solvation of polar groups, partly contrasting previous conclusions based on structural inspection. Our study illustrates that it is difficult and can be misleading to deduce binding mechanisms from structural models alone. Therefore, taking into account solvation effects as in MM-PBSA calculations is essential to elucidate molecular recognition.

1 Introduction

The anti-progesterone antibody DB3 shares 91% sequence identity to the Diels-Alderase antibody 1E9. However, 1E9 binds progesterone (PRG) and 5β-androstane-3,17-dione (5AD) weakly compared to DB3. Recently Piatesi *et al.*[1] have shown by site-directed mutagenesis and binding studies that only two mutations are required to interconvert the binding specificity of 1E9 and DB3. The $\text{Leu}^{\text{H47}}\text{Trp}/\text{Arg}^{\text{H100}}\text{Trp}$ 1E9 double mutant (1E9dm) binds both steroids with nanomolar (nM) affinity and achieves the binding specificity of DB3 for a panel of structurally and configurationally distinct steroid molecules. Recently the crystal structures of 1E9dm complexed with progesterone (PRG) or 5β-androstane-3,17-dione (5AD) have been determined by Verdino *et al*[2]. To elucidate the mechanism of the cross reactivity of 1E9 and 1E9dm with PRG and 5AD including solvent effects, we have conducted molecular mechanics–Poisson-Boltzmann surface area (MM-PBSA)[3] calculations studying the contributions arising from different molecular interactions to the binding affinities for different complexes.

2 Materials and Methods

The initial coordinates for our simulations were obtained from the X-ray crystallographic structures of the 1E9dm Fab complexed with progesterone (PDB ID: 2O5Y), and 5AD (PDB ID: 2O5Z) determined to 2.05 Å and 2.40 Å resolution, respectively[2]. The proteins were described using the Amber ff99SB force field. The ligands were assigned generalized amber force field (GAFF) atom types, and AM1-BCC atomic charges calculated with the *antechamber* module of Amber molecular dynamic package. Backward mutations were performed manually to study the binding of steroids to the wild type antibody 1E9. The

configurations were generated via simulations of the complexes in explicit water. A 10 ns simulation at constant pressure was conducted. Coordinates were saved after every 10 ps, resulting in 1000 configurations for each simulation. However, only 40 configurations were used to estimate the entropic contribution to the binding free energy.

2.1 MM-PBSA Calculations

The most common receptor-ligand (R-L) association reaction is governed by the following equation:

$$R + L \longleftrightarrow RL \tag{1}$$

where all the reactants are assumed to be in aqueous solution. The binding affinity is determined from the free energies of the receptor (R), the ligand (L), and the complex (RL):

$$\Delta G_{\text{bind}} = G_{\text{RL}} - (G_{\text{R}} + G_{\text{L}}) \tag{2}$$

The free energy of each species (R, L, RL) is estimated from

$$G = \langle E_{\text{MM}} \rangle + \langle G_{\text{pol}} \rangle + \langle G_{\text{np}} \rangle - T \langle S_{\text{MM}} \rangle. \tag{3}$$

Here, E_{MM} is the molecular mechanics gas-phase energy of the species, G_{pol} is the polar contribution to the solvation free energy of the species, estimated from the solution of the linear Poisson-Boltzmann (PB) equation ($G_{\text{pol,PB}}$), G_{np} is the non-polar solvation free energy, estimated from the solvent accessible surface area (SASA) of the species, T is the absolute temperature of the system, and S_{MM} is the entropy of the species, calculated from a normal-mode analysis of harmonic frequencies estimated at the molecular mechanics (MM) level. The gas-phase molecular mechanics energy E_{MM} can be expressed as

$$E_{\text{MM}} = E_{\text{cov}} + E_{\text{elec}} + E_{\text{vdW}}, \tag{4}$$

where E_{cov}, E_{elec}, and E_{vdW} denote the contributions from covalent, electrostatic, and van der Waals interactions, respectively. The contributions from nonpolar solvation, (G_{np}), was evaluated from

$$G_{\text{np}} = \gamma A_{\text{sasa}} \tag{5}$$

where A_{sasa} is the solvent accessible surface area (SASA) and $\gamma = 0.023 \, \text{kcal.mol}^{-1} \, \text{Å}^{-2}$. The SASA was estimated using a probe radius of 1.4 Å. The averages in Eq. 3 are calculated from an ensemble of molecular configurations taken from a molecular dynamics simulation to capture the effects of motion.

3 Results and Discussions

In order to understand the mechanisms underlying the binding of progesterone (PRG) and 5β-androstane-3,17-dione (5AD) to the antibodies 1E9 and 1E9dm, an energetic analysis using the MM-PB(GB)SA method was conducted. The structure of the 1E9dm-PRG complex showed an average RMSD of 1.46 Å, and an average RMSD of 1.56 Å was observed for the 1E9dm-5AD complex. The energetics of the binding of PRG and 5AD to 1E9 and 1E9dm obtained from the MM-PBSA calculations are shown in Tab. 1. Overall, PRG binds

more strongly to 1E9 and its variants than 5AD. In general, the largest contribution favoring binding is the van der Waals interaction between the binding partners, being in the range -40 kcal/mol to -43 kcal for PRG and -37 kcal/mol to -39 kcal/mol for 5AD. The nonpolar interactions with the solvent including the contribution from the hydrophobic effect make a contribution of -3.7 kcal/mol for PRG and in the range -3.5 kcal/mol to -3.6 kcal/mol for 5AD.

Contribution	PRG		5AD	
	WT	dm	WT	dm
ΔE_{elec}	-1.6(0.1)	-9.3(0.1)	-9.4(0.2)	-10.7(0.1)
ΔE_{vdW}	-39.5(0.2)	-42.7(0.1)	-36.8(0.2)	-38.6(0.1)
ΔG_{np}	-3.7(0.01)	-3.7(0.01)	-3.5(0.01)	-3.6(0.01)
ΔG_{pol}	22.5(0.2)	28.2(0.2)	28.2(0.3)	26.8(0.1)
$\Delta G_{\mathrm{solv}}^{(a)}$	18.8(0.2)	24.5(0.2)	24.7(0.3)	23.2(0.1)
$\Delta G_{\mathrm{pol,elec}}^{(b)}$	20.9(0.2)	18.9(0.2)	18.8(0.3)	16.1(0.1)
$-T\Delta S_{\mathrm{MM}}$	14.2(0.9)	15.4(1.2)	15.5(1.1)	15.0(0.8)
ΔG_{bind}	-8.1(0.9)	-11.7(1.2)	-6.0(1.1)	-11.1(0.8)

Table 1. Free energy terms (kcal/mol) for the binding of 1E9 variants to progesterone (PRG) and 5β-androstane-3,17-dione (5AD). Standard errors of the mean are given in parenthesis.

$^{(a)}$ $\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{pol}}$, $^{(b)}$ $\Delta G_{\mathrm{pol,elec}} = \Delta G_{\mathrm{pol}} + \Delta E_{\mathrm{elec}}$

To understand the difference in binding mechanisms between PRG-1E9dm and PRG-DB3, we have calculated the shift of binding free energies and their individual components for PRG-1E9dm with respect to PRG-DB3, by comparing our results for 1E9dm and the results from a study for DB3 by Peräkylä and Nordman[5], as shown in Tab. 2. For PRG-1E9dm the solutes remain more strongly exposed to the solvent than for PRG-DB3. On

Component	Value
$\Delta\Delta E_{\mathrm{elec}}$	+8.2(0.3)
$\Delta\Delta E_{\mathrm{vdW}}$	+4.8(0.3)
$\Delta\Delta G_{\mathrm{np}}$	+1.0(0.01)
$\Delta\Delta G_{\mathrm{pol}}$	-13.1(0.3)
$\Delta\Delta G_{\mathrm{solv}}$	-12.1(0.4)
$\Delta\Delta G_{\mathrm{pol,elec}}$	-4.9(0.4)
$\Delta\Delta G_{\mathrm{bind}}$	+1.3(0.2)

Table 2. Free energy terms of the PRG-1E9dm complex relative to the PRG-DB3 complex $^{(a,b)}$ in kcal/mol. Standard errors of the mean are given in parenthesis.

$^{(a)}$ $\Delta\Delta G = \Delta G^{\mathrm{1E9dm}} - \Delta G^{\mathrm{DB3}}$, $^{(b)}$ Free energy terms for the PRG-DB3 complex were obtained from Ref. 5.

the one hand, this leads to weaker van der Waals interactions between the binding partners and increased unfavorable nonpolar contributions to the binding free energy, stabilizing PRG-DB3 over PRG-1E9dm. On the other hand, the stronger solvation of polar groups stabilizes the PRG-1E9dm over the PRG-DB3 complex. Both effects largely compensate each other such that the net difference in affinity between the two complexes is small.

4 Concluding Remarks

Altogether, though the complexes formed by PRG or 5AD with 1E9dm and by PRG with DB3 have similar affinity, the binding mechanisms are different. Decreased van der Waals interactions observed for 5AD-1E9dm versus PRG-1E9dm or for PRG-1E9dm versus PRG-DB3 are energetically compensated by an increased solvation of polar groups. These findings do not confirm previous hypotheses on the underlying binding mechanism based on structural inspection.[2,4] This work illustrates that it is difficult and can be misleading to deduce binding mechanisms from structural models of the binding partners alone. In contrast, taking into account solvation effects as done here via MM-PBSA calculations is essential to understand molecular recognition.

Acknowledgments

References

1. A. Piatesi, C. Aldag, and D. Hilvert, *Switching antibody specificity through minimal mutation,* J. Mol. Biol. **377**, 993–1001, 2008.
2. P. Verdino, C. Aldag, D. Hilvert, and I. A. Wilson, *Closely related antibody receptors explot fundamentally different strategies for steroid recognition,* Proc. Nat. Acad. Sci. USA **105**, 11725–11730, 2008.
3. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, *Calculating structures andfree energies of complex molecules: combining molecular mechanics and continuum models,* Acc. Chem. Res. **33**, 889–897, 2000.
4. P. Kar, R. Lipowsky, and V. Knecht, *Importance of Polar Solvation for Cross-Reactivity of Antibody and Its Variants with Steroids*, J. Phys. Chem. B **115**, 7661–7669, 2011.
5. M. Peräkylä and N. Nordman, *Energetic analysis of binding of progesterone and 5β-androstane-3,17-dione to anti-progesterone antibody DB3 using molecular dynamics and free energy calculations,* Protein Eng. **14**, 753–758, 2001.

Resolving the Apparent Gap in Complexity between Simulated and Measured Kinetics of Biomolecules

Bettina G. Keller, Jan-Hendrik Prinz, and Frank Noé

Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
E-mail: {bettina.keller, jan-hendrik.prinz, frank.noe}@fu-berlin.de

Molecular simulations of biomolecules often reveal a complex picture of the their kinetics, whereas kinetic experiments typically seem to indicate considerably simpler two- or three-state kinetics. Markov state models (MSM) provide a tool to link between simulation and experiment, and to resolve this apparent contradiction.

1   Introduction

Molecular simulations of large biomolecules typically reveal a complex picture of the free-energy surface with many kinetically relevant states[1–3]. Complementary to this, advanced experimental techniques allow probing the equilibrium kinetics of biomolecules directly. This can either be done by perturbation techniques, such as temperature or pressure jump, or by equilibrium experiments in which the measured signal is autocorrelated to obtain kinetic information, as e.g. in fluorescence correlation spectroscopy. These measurements can be described by dynamical fingerprints, i.e. densities of relaxation timescales where each peak corresponds to an exponential relaxation process. In many cases, single- or double-peaked fingerprints are found, suggesting that a two- or three-state model may provide a satisfactory description of the biomolecule studied[4,5].

We sketch an approach combining Markov state models (MSM)[3,6] of the simulated dynamics with dynamical fingerprints, which allows addressing the following questions: (i) Is the largest relaxation timescales observed always due to the folding process? (ii) Can a given experiment detect all relaxation processes that are present in the dynamics of the molecule? (iii) Are the processes observed in perturbation experiments the same as those observed in equilibrium experiments? We illustrate our findings using a four-state model of a protein folding equilibrium.

2   Theory

In MSMs of conformational kinetics, the conformational space of the molecule is discretized into N states. The kinetics between the states is described by transition probability t_{ij} of going from a state i to a state j within a time step τ, which are summarized in a transition matrix $\mathbf{T}(\tau)$. The entire information of the kinetics of the system is contained in this matrix. In particular, each (left) eigenvector $\mathbf{l}_i$ of the transition matrix represents one of the kinetic processes of the system. For a detailed review of the theory of MSM and their application to the simulation of biomolecules see Ref. 6 and 3.

Kinetic experiments yield time series $\mu(t)$, either of the observed signal a directly (perturbation experiments), or of the autocorrelation of the signal (equilibrium experiments).
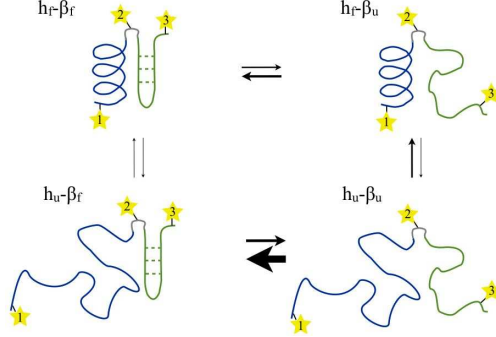
Figure 1. Sketch of a protein folding equilibrium. The arrows represent possible transitions between conformational states. Their thickness corresponds to the transition probability. The yellow stars represent possible chromophore attachment points.

From physical principles, we expect that in both cases $\mu(t)$ is a noisy realization of a sum of multiple exponential functions

$$\mu(t) = \int_{t'} dt' \gamma(t') \exp\left(-\frac{t}{t'}\right) \approx \sum_i^N \gamma_i \exp\left(-\frac{t}{t_i'}\right).$$ (1)

$\gamma(t')$ is inverse the Laplace transform of $\mu(t)$ and is called the *dynamical fingerprint*[7]. The approximation in Eq. 1 results from the fact that any experimental signal is time-discretized, and that in an MSM-representation of the kinetics the number of processes which contribute to the dynamical fingerprint is limited to the number of states N. The relaxation timescales t_i of the system are linked to the eigenvalues λ_i of the transition matrix by $t_i' = -\tau / \ln \lambda_i$ One can derive analytical expressions[7,8] for the amplitudes γ_i of equilibrium experiments $\gamma_i^{\boldsymbol{\pi},a} = \langle \mathbf{a}, \mathbf{l}_i \rangle^2$ and of perturbation experiments $\gamma_i^{\mathbf{p}(0),a} = \langle \mathbf{a}, \mathbf{l}_i \rangle \langle \mathbf{p}'(0), \mathbf{l}_i \rangle$. $\mathbf{p}'(0)$ is the excess probability density with $p_j'(0) = p_j(0)/\pi_j$, where $\boldsymbol{\pi}$ is the equilibrium density. $\mathbf{a}$ is the observable vector which associates each states in the MSM with a signal strength of the observable a.

3 Model System and Results

Each of the secondary structure elements of the protein folding model in Fig. 1 can fold and unfold in a single distinct step. This leads to a MSM of the conformational equilibrium with $N = 4$ metastable states. The transition probabilities are represented by the thickness of the arrows in Fig. 1. None of the eigenvectors of the MSM (Fig. 2.a) reflects our notion of folding. Hence, the folding process may not be contained in a kinetic process and can thus not always be associated with a single folding rate.

The yellow stars in Fig. 1 mark possible chromophore attachment points. We choose our observable vectors to resemble FRET constructs, i.e. two chromophores attached to the molecule with strong signal if they are close to each other, weak signal if they are further apart (Fig. 2.c).
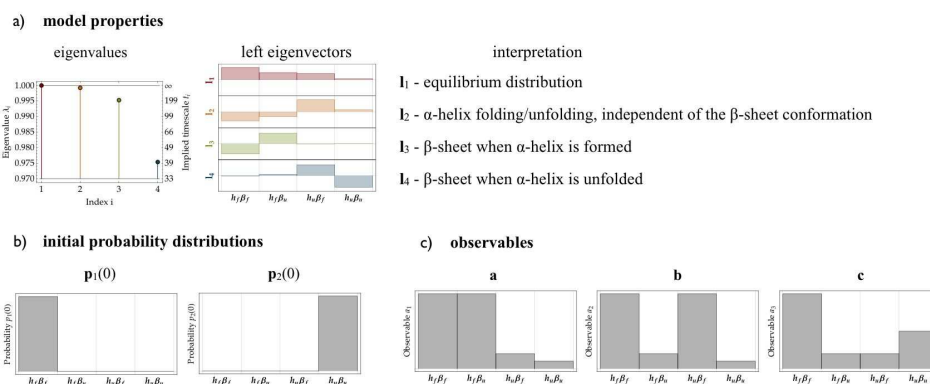
Figure 2. Markov model and experimental setup for the protein folding model.
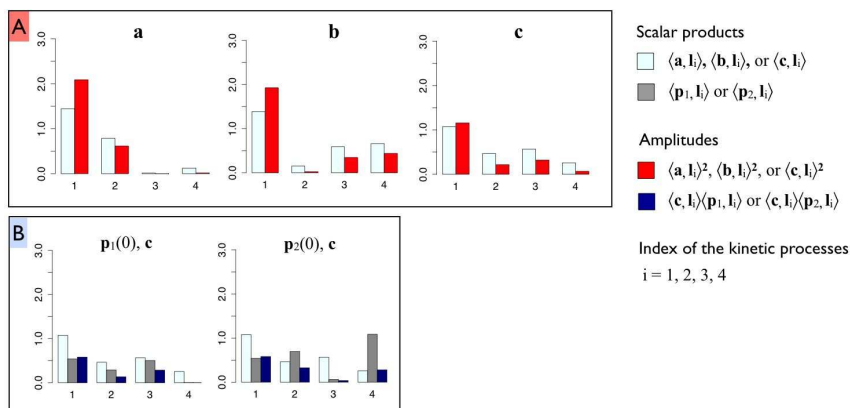
Figure 3. Amplitudes of the dynamical fingerprints of a variety of experimental setups for the protein folding model. (A) Equilibrium experiment, dynamical fingerprint of the autocorrelation function of observables $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. (B) Perturbation experiment, dynamical fingerprint of the decay signal of the observable $\mathbf{c}$ combined with initial distributions $\mathbf{p}_1(0)$ and $\mathbf{p}_2(0)$, respectively.

Fig. 3.A show the dynamical-fingerprint amplitudes of equilibrium experiments using each of the three possible observable vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. The scalar product of observable vector $\mathbf{a}$ with eigenvectors $\mathbf{l}_3$ and $\mathbf{l}_4$, is close to zero and therefore the corresponding experiment is insensitive to these processes. This is in line with the intuition that an experiment in which the chromophores are attached at site 1 and 2 will be most sensitive to conformational changes of the α-helix. Conversely, observable $\mathbf{b}$ has little overlap with eigenvectors $\mathbf{l}_2$, and the corresponding experiment is insensitive to this kinetic process representing the α-helix folding. Only with observable $\mathbf{c}$ all relevant kinetic processes can be observed. This shows that a single experiment will typically be insensitive to some of the processes present in the system and explains why kinetics appear often simpler in experiments than in simulations.

63

A perturbation experiment cannot reveal a kinetic process which would be invisible in an equilibrium experiment with the same observable. This is easily seen by considering that the amplitude is given as $\gamma_i^{\mathbf{p}(0),a} = \langle \mathbf{a}, \mathbf{l}_i \rangle \langle \mathbf{p}'(0), \mathbf{l}_i \rangle$. If the first factor is zero, i.e. the process is not detectable in an equilibirum experiment, the corresponding amplitude in the fingerprint of the perturbation experiment will also be zero - independent of the initial distribution $\mathbf{p}(0)$. However, not all kinetic processes are involved in relaxing a particular initial distribution to the equilibrium distribution and therefore the second factor can become zero, too. This is shown in Fig. 3.B, where $\langle \mathbf{p}'_1(0), \mathbf{l}_3 \rangle \approx 0$ and $\langle \mathbf{p}'_2(0), \mathbf{l}_4 \rangle \approx 0$, respectively (see also Fig. 2.a and see Fig. 2.c). Consequently, these perturbation experiments are insensitive to the third and fourth kinetic process respectively, even though they are conducted with an observable which sensitive to all processes in the system.

Given a MSM of the conformational kinetics of biomolecule, our approach can also be used to suggest optimal attachment points for the chromophores[7,8].

Acknowledgments

References

1. S. Muff and A. Caflisch, *Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein*, Proteins-Struct., Funct., Bioinf., **70**, no. 4, 1185–1195, 2008.

2. F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, *Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations*, Proc. Natl. Acad. Sci. USA, **106**, no. 45, 19011–19016, 2009.

3. G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, *Progress and challenges in the automated construction of Markov state models for full protein systems*, J. Chem. Phys., **131**, no. 12, 124101, 2009.

4. Marcus Jäger, Houbi Nguyen, Jason C. Crane, Jeffery W. Kelly, and Martin Gruebele, *The folding mechanism of a beta-sheet: the WW domain*, J. Mol. Biol., **311**, no. 2, 373–393, 2001.

5. Hannes Neuweiler, Marc Löllmann, Sören Doose, and M. Sauer, *Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy*, J. Mol. Biol., **365**, 856–869, 2007.

6. J-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Fischbach, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *Markov model and molecular kinetics: generation and validation*, J. Chem. Phys., **134**, 174105, 2011.

7. F. Noé, S. Doose, I. Daidone, M. Lollmann, M. Sauer, J. D. Chodera, and J. C. Smith, *Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments*, Proc. Natl. Acad. Sci. USA, **108**, no. 12, 4822–4827, 2011.

8. B. G. Keller, J.-H. Prinz, and F. Noé, *Markov models and dynamical fingerprints: unraveling the complexity of molecular kinetics.*, Chem. Phys., submitted 2011.

Thiopeptide Antibiotics and the Ribosomal 23S·L11 Subunit – A Challenging Use Case for Semi-automatic Force-field Development

Antje Wolf, Dirk Reith, and Karl N. Kirschner

Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)
Department of Computational Chemistry Engineering
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
E-mail: karl.kirschner@scai.fraunhofer.de

Force-field optimization remains an important endeavor for modeling chemically diverse systems at atomistic resolution. It also remains a time-consuming multiscale process involving multiple steps and data transferal. One goal of our research is to develop accurate and reliable parameters in a reasonable time and as error free as possible. Thus, we have designed two modular scientific workflows, Wolf₂ Pack and GROW, that incorporate our beliefs for how quantum mechanical (QM) and experimentally gained knowledge should be transferred to molecular mechanics (MM) models.

Herein, we have applied Wolf₂ Pack to develop parameters for thiopeptide antibiotics, natural products that bind to a ribosomal protein-RNA subdomain. For illustration we present the performance of this force field when used in a molecular dynamics (MD) simulation of thiostrepton (thio) immersed in an organic solvent mixture.

1   Introduction

1.1   Force Fields

Molecular force fields have a rich ~60-year history, with the theory's foundation coming from 1930s vibrational spectroscopy. However, it wasn't until the 1970s that their utility and popular use as a research tool took a strong hold. Today, force fields continue to be an important part of computational research, and there exists a variety of published parameter sets,[a] whose implementations range from atomistic to mesoscale investigations.

Force-field development is a time consuming process involving multiple steps and data transferal, and is the basis of any meaningful application-type study using MM theory. To assist us in developing reliable molecular models, we are developing two scientific workflows for force-field optimization (Wolf₂ Pack[2] and GROW[3]) that incorporate our beliefs for transferring QM and experimentally gained knowledge to MM models. GROW's purpose is to optimize nonbonded parameters for reproducing target experimental data via MD simulations. Wolf₂ Pack's purpose is to optimize intramolecular parameters using QM data, by allowing the user to steer the optimization process.[b] An important part of Wolf₂ Pack is that parameters are developed in a systematic and consistent fashion. This is reflected by the generation of a QM molecular database, which currently contains 335 molecules and 1237 potential energy curves for various degrees of freedom.

---

[a] Please see Ref. 1 for a short summary.

[b] Both Wolf₂ Pack and GROW are able to optimize partial atomic charges but using different approaches.

The bio-computational chemist community primarily *uses* force fields in their research investigations, with a small proportion of groups developing them. Thus, most of the simulations are on "traditional" systems involving standard amino and nucleic acids, where the force fields are well established. By providing users with a robust parametrization tool, we hope to increase the availability of reliable parameters for modeling new diverse biomolecular (e.g. lipids and natural products) and chemical (e.g. ionic liquids) systems.

1.2 Thiopeptides

Naturally occurring thiopeptides, such as thio[4,5], inhibit bacterial protein biosynthesis by binding to a cleft formed by the L11·23S ribosomal subunit. Currently the compounds are used only in the treatment of animals. Understanding their molecular properties and activity better, one can work towards designing a novel compound that can be used in human therapy. Experimentalists have explored the structure of these compounds and their binding using X-ray, NMR, and chemical modification approaches[6–8]. Modeling these natural products is challenging for computational chemists using force-field based methods; like many natural products they have chemical functionalities and bonding that are not present in traditional protein, DNA, and RNA macromolecules. Thus, an alternative force field is needed whose parameters cover the unique functionalities within thiopeptides.

Herein we briefly discuss the building and optimization of our model for thio. The Wolf$_2$ Pack workflow allowed us to optimize all of thio's internal degrees of freedom, resulting in a robust force field for modeling thiopeptides. Finally, structural results of MD simulations involving thio immersed in an organic solvent are compared to an NMR-obtained structure.

2   Methodology

Thio was separated into its constituent residues as defined by Bond and coworkers (Fig. 1)[9]; each non-amino acid residue was capped using the nearest functional group, and terminated using a methyl group. The residues were optimized using HF/6-31G(d) theory level, and electrostatic potentials were subsequently computed using CHELPG. A RESP[10] fit was performed via R.E.D.[11] using a 0.01 weighting factor to obtain charges compatible with Parm99SB. Initial parameters for the standard amino acid residues were taken from Parm99SB[12], while the non-standard residue parameters were taken from GAFF[13]. Using Wolf$_2$ Pack[2] a second parameter set was optimized for modeling thio, chloroform, and methanol. Since we plan to model thio bound to the ribosome, we chose to use Parm99SB Lennard-Jones parameters instead of optimizing them using GROW. For further methodological information, please see Ref. 2.

Both parameter sets were used in separate 20 ns isothermal-isobaric MD simulations performed at 300.15 K. 1-4 electrostatic and nonbonded scaling factors were set to one during the parametrization and MD simulations. All MM-based and QM calculations were performed using AMBER[14] and GAMESS[15].

3   Results and Discussion

Our initial MD investigation was done using the Parm99SB/GAFF parameter combination. Using such a combined parameter set is a plausible route when confronted with an incom-
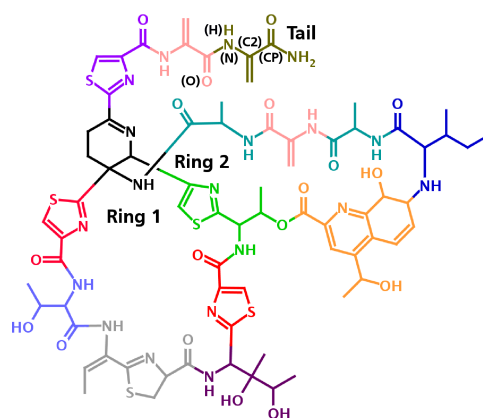
Figure 1. Thiostrepton structure color coded according to residues[9]. Key atom types within the tail are shown in parenthesis.

plete force field. However due to the developmental differences between force fields, such a "Frankenstein" approach can lead to untrustworthy results. Ultimately we wanted a force field that was more reliable and whose errors could be tracked down and corrected more easily. Thus, we used $Wolf_2$ Pack to optimize thio's intramolecular degrees of freedom. In total, 35 bonds, 89 angles, and 153 torsions, using 18 atom types, were optimized. Over 500 potential energy curves ($>$10,000 QM data points) were used in the parameter training and transferability testing.

To partially validate the parameters, we created a model to reproduce the condition of the thio NMR experiment[8], by immersing thio in a 5:1 chloroform:methanol mixture. Two MD simulations were performed using the Parm99SB/GAFF and $Wolf_2$ Pack optimized force fields. While the Parm99SB/GAFF parameter set performs adequately, as seen in Figures 2a and 2b the optimized force field reproduces the NMR structure better. However, a close examination of the torsion values and dynamics within thio's tail residues revealed some questionable behavior by the optimized force field. We reexamined the tail's torsion

Angle	$Wolf_2$ Pack Initially Optimized				$Wolf_2$ Pack Reoptimized			
	V_1	V_2	V_3	V_4	V_1	V_2	V_3	V_4
C2-C2-CP-N	-0.25	-1.56	0.00	0.00	-0.25	-1.56	0.00	0.00
C2-C2-CP-O	0.25	-1.56	0.00	0.00	0.25	-1.56	0.00	0.00
N-C2-CP-N	1.00	-2.00	0.00	0.00	0.50	-1.20	0.10	0.00
N-C2-CP-O	0.00	0.00	0.00	0.00	-0.50	-1.20	-0.10	0.00
C2-C2-N-CP	0.00	-0.88	0.00	0.00	0.00	-1.00	0.00	0.00
C2-C2-N-H	0.00	-0.88	0.00	0.00	0.00	-1.00	0.00	0.00
CP-C2-N-CP	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.15
CP-C2-N-H	0.00	0.00	0.00	0.00	-0.25	0.00	0.00	0.15

Table 1. Key coupled torsion parameters that were optimized for modeling thio's tail (Fig. 1).

67

parameters and found an alternative solution for coupled parameters describing the rotation about C2-CP and C2-N bonds (see Tab. 1). A third MD simulation was performed and provided improved structural agreement to the NMR data (Fig. 2c). Note that we also obtain better agreement in the Ring 2 structure, which contains a residue that is dependent upon these reoptimized parameters (Fig. 1).

(a)

(b)

(c)
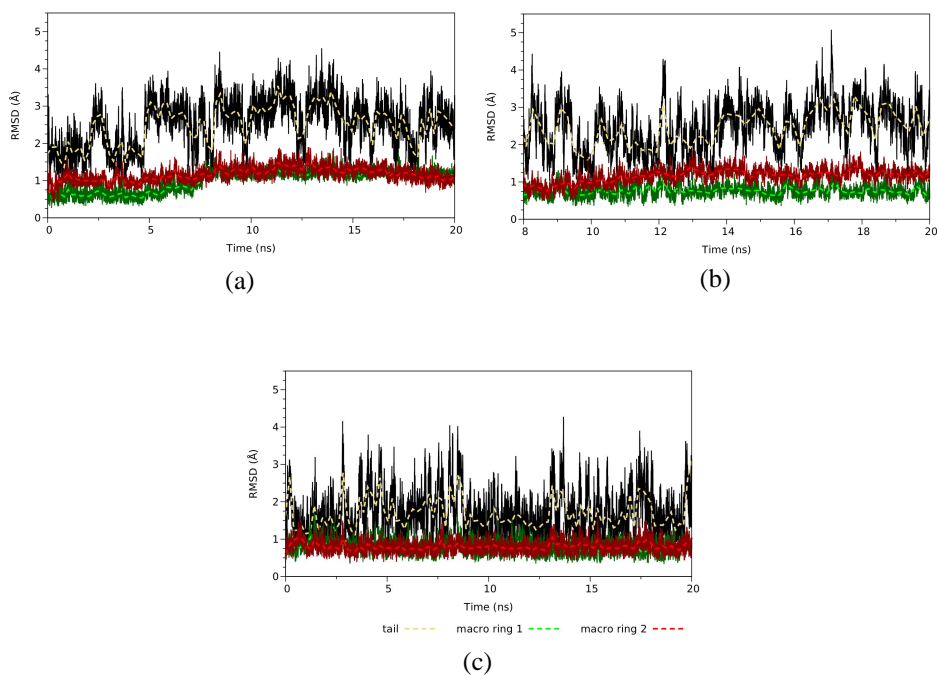
Figure 2. Root-mean-squared deviation of the thiostrepton model with respect to the NMR structure versus time, using the following force fields: (a) the Parm99SB/GAFF combination, (b) the initial $Wolf_2$ Pack optimized, and (c) the $Wolf_2$ Pack reoptimized.

4   Conclusion

In our pursuit to develop reliable force fields we have designed two modular scientific workflows, $Wolf_2$ Pack and GROW, that optimize intra- and intermolecular parameters. Herein we report the utilization of $Wolf_2$ Pack to develop parameters for thio, a thiopeptide antibiotic that binds to the ribosome. Thiopeptides are challenging for force fields and their development due to their large sizes and diverse chemical composition. By using $Wolf_2$ Pack, one can efficiently generate the QM target data, develop the partial atomic changes, and steer the parameter optimization, as the user deems appropriate, for their model. As shown with thio, the user can systematically trace force-field errors back to individual parameters, which can be adjusted to improve MM and MD results.

References

1. B. Hess, C. Kutzner, D. van der Spool and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput. **4**, 435–447, 2008.

2. D. Reith and K. N. Kirschner, *A Modern Workflow for Force-Field Development - Bridging Quantum Mechanics and Atomistic Computational Models*, Comput. Phys. Commun. **182**, 2184–2191, 2011.

3. M. Hülsmann and T. Köddermann and J. Vrabec and D. Reith, *GROW: A Gradient-Based Optimization Workflow for the Automated Development of Molecular Models* Comput. Phys. Commun. **181**, 499-513, 2010.

4. R. Donovick, J. F. Pagano, H. A. Stout and M. J. Weinstein, *Thiostrepton, a New Antibiotic. I. in vitro Studies.*, Antibiot. Annu. **3**, 554–559, 1955.

5. J. D. Dutcher and J. Vandeputte, *Thiostrepton, a New Antibiotic. II. Isolation and Chemical Characterization.*, Antibiot. Annu. **3**, 560–561, 1955.

6. H. R. A. Jonker, S. Baumann, A. Wolf, S. Schoof, F. Hiller, K. W. Schulte, K. N. Kirschner, H. Schwalbe and H.-D. Arndt, *NMR Structures of Thiostrepton Derivatives for Characterization of the Ribosomal Binding Site*, Angew. Chem. Int. Ed. **50**, 3308–3312, 2011.

7. J. M. Harms, D. N. Wilson, F. Schluenzen, S. R. Connell, T. Stachelhaus, Z. Zaborowska, C. M. T. Spahn and P. Fucini, *Translational Regulation Via L11: Molecular Switches On the Ribosome Turned On and Off by Thiostrepton and Micrococcin.*, Mol. Cell **30**, 26–38, 2008.

8. H. R. A. Jonker, S. Ilin, S. K. Grimm, J. Wohnert and H. Schwalbe, *L11 Domain Rearrangement Upon Binding to RNA and Thiostrepton Studied by NMR Spectroscopy*, Nucleic Acids Res. **35**, 441–454, 2007.

9. C. S. Bond, M. P. Shaw, M. S. Alphey and W. N. Hunter, *Structure of the Macrocycle Thiostrepton Solved Using the Anomalous Dispersion Contribution of Sulfur.*, Acta Crystallogr. D **57**, 755–758, 2001.

10. J. Wang, P. Cieplak and P. Kollman, *How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?*, J. Comput. Chem. **21**, 1049–1074, 2000.

11. F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building*, Phys. Chem. Chem. Phys. **12**, 7821–7839, 2010.

12. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters*, Proteins **65**, 712–725, 2006.

13. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Development and Testing of a General Amber Force Field*, J. Comput. Chem. **25**, 1157–1174, 2004.

14. D. A. Case et. al., *AMBER* 9, 2006.

15. M. S. Gordon and M. W. Schmidt, *Advances in Electronic Structure Theory: GAMESS a decade later, in Theory and Applications of Computational Chemistry: the First Forty Years* M. S. Gordon, M. W. Schmidt, and C. E. Dykstra, eds., Elsevier, Amsterdam, Boston, 1167–1189, 2005.

Influence of Thermostabilizing Mutations in Fructose-1,6-bisphosphate Aldolases Revealed by Constraint Network Analysis

Doris L. Klein[1], Jijun Hao[2], Alan Berry[2], and Holger Gohlke[1]

[1] Department of Mathematics and Natural Sciences, Heinrich-Heine-University
40225 Düsseldorf, Germany
E-mail: gohlke@uni-duesseldorf.de

[2] The Astbury Centre for Structural Molecular Biology, University of Leeds
Leeds LS2 9JT, United Kingdom

Four fructose-1,6-bisphosphate aldolases obtained by directed evolution show significant differences in the thermostability, although the differences in sequence and structure are only minor. This raises the question as to what is the molecular origin of the thermostabilization. Here, we apply Constraint Network Analysis (CNA) to link the mechanical rigidity of a protein structure to its thermostability. The analyses reveal structural weak spots of the aldolases. Mutations that lie within these weak spots are found to reinforce the network of non-covalent interactions and, thus, increase the protein's thermostability.

1 Introduction

Most proteins have efficiently adapted towards the conditions of their particular environmental niche[1]. One major requirement is the adaption towards temperature. Proteins are usually stable and active under a wide temperature range, which is made possible by versatile mechanisms[2]. Still, there is a need of adapting proteins towards specific conditions of industrial processes. Fructose-1,6-bisphosphate aldolase[3] (FBP-aldolase) (Fig. 1) has been restricted for use in synthetic organic chemistry due to its low thermostability. The enzyme belongs to the $(\alpha/\beta)_8$ barrel family[4] and cleaves fructose-1,6-bisphosphate into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate[5]. By means of directed evolution, Hao *et al.* engineered aldolases with enhanced thermostability[3]. The identified aldolases 1-44F2, 2-15B2, 3-4C10, and 4-43D6, henceforth referred to as A1, A2, A3, and A4, respectively, show only minor differences in sequence and structure (see Tab. 1) but display significant differences in the temperature where 50 percent of the proteins are still active (T_{50}) of up to 9.3 K[3].This raises the question as to what is the molecular origin of the thermostabilization of these enzymes.

2 Methods

The program Constraint Network Analysis (CNA)[7,8] was used to analyze A1-A4 in terms of structural determinants of thermostability. Initially, the protein is modeled as a constraint network, where atoms are represented as vertices (joints) and covalent and non-covalent bonds are represented as edges (constraints). The mechanical rigidity (structural stability) of proteins is largely determined by non-covalent interactions, i.e. hydrogen bonds, salt
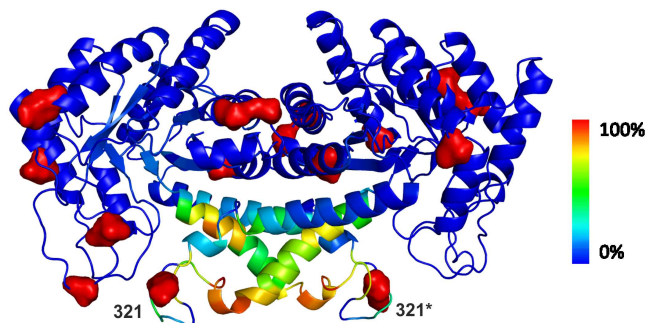
Figure 1. Quaternary structure of FBP-aldolase A2. The weak spot probability is color-coded from dark blue (0%) to red (100%). Mutations from aldolase mutant A1 to A2 are shown in red surface representation.

	$T_{50}^{[b]}$	A1	A2	A3	A4
A1	52.2	-	0.25	0.19	0.19
A2	57.6	97.8	-	0.25	0.21
A3	59.3	97.6	97.8	-	0.13
A4	61.7	96.5	97.8	98.3	-

Table 1. T_{50} values, all atom RMSD, and sequence identities of FBP-aldolase mutants A1-A4[a].

[a] The all atom RMSD (in Å) is given above the empty diagonal, the sequence identity (in %) below the empty diagonal.
[b] In K.

bridges, and hydrophobic interactions[9]. Here, hydrogen bonds are ranked according to an energy function, and only hydrogen bonds with an energy of lower than -0.1 kcal mol^{-1} are initially retained in the network. Next, the constraint network is analyzed with the software FIRST in order to determine the local network rigidity.

In order to link the mechanical rigidity of the protein structure to its thermostability, the thermal unfolding of the protein is simulated by gradually removing hydrogen bonds from the network in the order of increasing strength[7,8]. This is based on the idea that stronger hydrogen bonds will break at higher temperatures than weaker ones. During the thermal unfolding, the protein develops from a largely rigid structure at low temperatures to a mostly flexible one at high temperatures. When analyzing macroscopic properties of the network during the thermal unfolding by indices from percolation theory[7,8], two phase transitions are observed (Fig. 2): The first one describes the break-down of the completely rigid network into a number of rigid clusters, while the second one describes the loss of the residual rigidity and, hence, the transition to a network that is completely flexible. The second transition corresponds to the folded-unfolded transition in experimental protein unfolding[7]. Thus, the phase transition temperature (T_p) associated with it can be related to the experimental melting temperature or, if this is not available, to T_{50}. Here, T_p is aver-
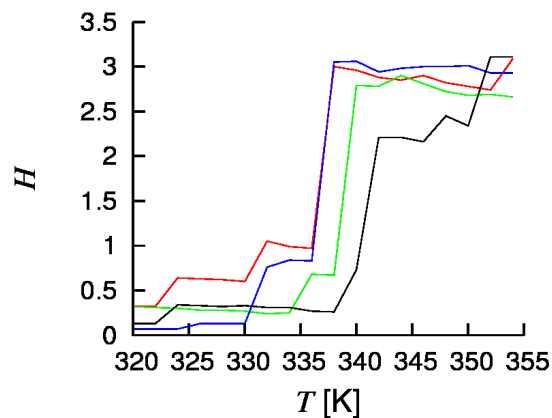
Figure 2. Cluster configuration entropy curves of aldolases A1 (red), A2 (green), A3 (blue), and A4 (black) showing the thermal unfolding of the protein structures. The onset of the largest change in H determines T_p.

aged over an ensemble of 125 conformations generated by all-atom molecular dynamics simulation of 5 ns length.

3   Results and Discussion

Differences in the computed phase transition temperatures (ΔT_p) of the four aldolase systems in comparison to the ΔT_{50}-values are shown in Fig. 3. From a qualitative point of view, all three mutants A2, A3, and A4 are computed to have a higher phase transition temperature than A1, and A4 is correctly identified to be more thermostable than A2. In contrast, the computed thermostability of A3 is too low compared to A2. Still, we note that $\Delta T_{50,A3-A2}$ amounts to only 1.7 K, which is close to the uncertainty of the computations

Figure 3. Experimental ΔT_{50} (blue) and computed ΔT_p (orange) for mutants A1-A4. The T_{50} and T_p values of A1 are used as a reference point.

of ΔT_p of 1.2 K. From a quantitative point of view, all computed changes in the thermostability fall short of the changes determined by experiment. However, this will not influence any further analyses because the absolute magnitude of the phase transition temperature depends on an empirical scaling[8], which is identical for all four systems.
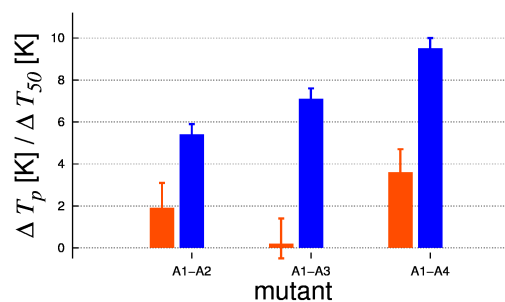
Next, going from A1 to A2, two mutations at positions 321 in chain a and 321* in chain b were investigated in more detail in order to understand the mechanisms of protein structural adaptation that lead to higher thermostability. Importantly, these mutations lie in weak spot regions determined by CNA (Fig. 1). These regions are considered the weakest part of the proteins unfolding nucleus, and previous analyses have revealed that thermostabilizing mutations are often located in such weak spots[7]. Indeed, molecular dynamics simulations show that Met321_{A1} only forms one hydrogen bond between its main chain oxygen and the main chain nitrogen of Gly314_{A1}. Lys321_{A2} shows the same interaction with Gly314_{A2}, but additionally forms a hydrogen bond with its guanidine moiety either to Glu317_{A2} or to Asp320_{A2}. This reinforces the network of non-covalent interactions and, thus, increases the protein's thermal stability. Similarly, Met$321*_{A1}$ is stabilized by just one main chain hydrogen bond, while the equivalent residue Lys$321*_{A2}$ forms an additional hydrogen bond with its side chain and, therefore, contributes to the higher overall stability of A2 compared to A1, too. Overall, these initial analysis demonstrate that CNA can help understanding the influence of thermostabilizing mutations by pointing to regions where such mutations excert a pronounced effect.

References

1. Ness, J. E., Del Cardayre, S. B., Minshull, J., and Stemmer, W. P. *Molecular breeding: the natural approach to protein design.*, Adv Protein Chem **55**, 261-292, 2000.
2. Vogt, G., Woell, S., and Argos, P. *Protein thermal stability, hydrogen bonds, and ion pairs.*, Adv Protein Chem **269**, 631-643, 1997.
3. Hao, J., and Berry, A. *A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents.*, Protein Eng **17**, 689-697, 2004.
4. Sygusch, J., Beaudry, D., and Allaire, M. *Molecular Architecture of Rabbit Skeletal-Muscle Aldolase at 2.7-a Resolution.*, Proc Natl Acad Sci USA **84**, 7846-7850, 1987.
5. Zgiby, S. M., Thomson, G. J., Qamar, S., and Berry, A. *Exploring substrate binding and discrimination in fructose1,6-bisphosphate and tagatose 1,6-bisphosphate aldolases.*, Eur J Biochem **267**, 1858-1868, 2000.
6. Ramsaywak, P. C., Labbe, G., Siemann, S., Dmitrienko, G. I., and Guillemette, J. G. *Molecular cloning, expression, purification, and characterization of fructose 1,6-bisphosphate aldolase from Mycobacterium tuberculosis-a novel Class II A tetramer.*, Protein Expres Purif **37**, 220-228, 2004.
7. Radestock, S., and Gohlke, H. *Exploiting the link between protein rigidity and thermostability for data-driven protein engineering.*, Eng Life Sci **8**, 657-657, 2008.
8. Radestock, S., and Gohlke, H. *Protein rigidity and thermophilic adaptation.*, Proteins **9**, 1089-1108, 2011.
9. Gohlke, H., Kuhn, L. A., and Case, D. A. *Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach.*, Proteins **56**, 322-337, 2004.

Achieving Numerical Stability in Analytical Computation of the Molecular Surface and Volume

Konstantin Klenin[1], Frank Tristram[2], Timo Strunk[2], and Wolfgang Wenzel[2,3]

[1] Steinbuch Center for Computing, Karlsruhe Institute of Technology
P.O. Box 3640, D-76021 Karlsruhe, Germany
E-mail: konstantin.klenin@kit.edu

[2] Institute of Nanotechnology, Karlsruhe Institute of Technology
P.O. Box 3640, D-76021 Karlsruhe, Germany
E-mail: {frank.tristram, timo.strunk, wolfgang.wenzel}@kit.edu

[3] Center for Functional Nanostructures, Karlsruhe Institute of Technology
P.O. Box 6980, D-76049 Karlsruhe, Germany

The computational effort of biomolecular simulations can be significantly reduced by means of implicit solvent models in which the energy contains a term depending on the surface area or/and volume of the molecule. Although the analytical formulas for these quantities are known, their practical implementation in a computer program is not trivial because of arising numerical instabilities. We propose a new approach to solution of this problem. Our approach has been implemented in a C++ header-only template library, called POWERSASA, for computation of the molecular surface, volume and their derivatives with respect to atomic coordinates.

1 Introduction

In the last decade, great progress has been made in all-atom computer simulations of biological macromolecules, such as proteins and nucleic acids. In particular, force fields were optimized to provide better computational efficiency. One attractive choice is the use of an implicit solvent model, which permits elimination of many degrees of freedom associated with the surrounding water molecules and ions. In this model, the simulated molecule is treated as a union of balls, so that each such ball corresponds to an atom. The ball radius is usually taken larger than the standard van der Waals value by the effective radius of the water molecule (~ 1.4 Å)[1]. In many currently used force fields, the effective solvation energy depends on the surface area and/or the volume of this geometrical object[2,3]. For some simulation methods, such as Monte Carlo, it is enough to be able to evaluate the energy as a function of the atomic coordinates. For others, like molecular dynamics, evaluation of the forces acting on the atoms is required. Hence, it is important to have efficient algorithms for computation not only of the surface area and volume but also of the derivatives of these quantities with the respect to atomic coordinates as well.

2 Theory

Consider a molecule as a union of overlapping balls (atoms) of different radii. Let us select a single arbitrary atom. We denote its radius by ρ and the position of its center by $\mathbf{r}$. The buried area of its surface has a boundary that consists of the arcs resulting from the
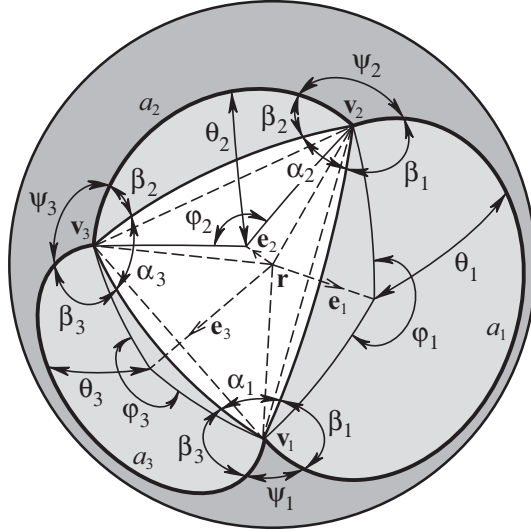
Figure 1. The surface of the selected atom (example). The buried area is shown in white and light-gray. The arcs forming the boundary are represented by the thick solid lines. The thinner solid lines (without arrows) are geodesics. The dashed lines are straight segments that do not belong to the surface. (The figure is reprinted from Ref. 4. Copyright 2011, Wiley Periodicals, Inc.)

intersection with the neighboring atoms. Consider a single contour formed by n such arcs as illustrated in Fig. 1. We denote the arcs by $a_1, a_2, \ldots, a_n$, with the indices increasing as one moves round the buried area in an anticlockwise manner. For simplicity, we ascribe the same indices to the corresponding neighboring atoms (so that, in general, one atom may obtain several different indices). The accessible surface area associated with the contour is given by the Connolly formula[5]:

$$S = \left[2\pi + \sum_{i=1}^{n}(\varphi_i \cos\theta_i + \psi_i - \pi)\right]\rho^2, \tag{1}$$

where the meanings of the angles φ_i, ψ_i, and θ_i are illustrated in Fig. 1. In the general case, when a boundary of the buried area consists of several contours, each such contour provides an additive contribution to the total accessible area, which, however, should be put into the interval $(0, 4\pi\rho^2)$ by subtracting $4\pi\rho^2$ the necessary number of times. The accessible area derivative with respect to the position $\mathbf{r}_i$ of the center of the ith neighboring atom is known to be[4]:

$$\frac{dS}{d\mathbf{r}_i} = \frac{\varphi_i \rho}{2}\left(1 + \frac{\rho^2 - \rho_i^2}{b_i^2}\right)\mathbf{e}_i - \frac{\rho}{b_i}[(\mathbf{v}_{i+1} - \mathbf{v}_i) \times \mathbf{e}_i], \tag{2}$$

where ρ_i is the radius of the ith neighbor, $b_i = |\mathbf{r}_i - \mathbf{r}|$ is the distance to it, $\mathbf{e}_i = (\mathbf{r}_i - \mathbf{r})/b_i$ is the unit vector pointing in the direction from $\mathbf{r}$ to $\mathbf{r}_i$, and $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ are the end points of the arc a_i. The analytical expressions for the volume and its derivatives are also available[4].

From the computational point of view, the most difficult problem is to construct the contours of arcs and, in particular, to identify the contour vertices $\mathbf{v}_i$ among all the other

triple intersections of the atomic spheres. It can be efficiently done by means of the so called power diagram[6,7] that is defines as follows.

The power distance from an arbitrary point $\mathbf{x}$ in the three-dimensional space to atom i is defined as $(\mathbf{x} - \mathbf{r}_i)^2 - \rho_i^2$. The power cell of atom i is the set of all points, for which the power distance to atom i is not larger than to any other atom. The collection of all the cell defines the power diagram of the given conformation of a molecule. The contour vertices $\mathbf{v}_i$ can be identified by the following two properties: (1) they lie on the edges of the power diagram, and (2) they have zero power distance to the atoms, the cells of which share the corresponding edge.

3   Numerical Stability

In Ref. 4 we proposed a recursive algorithm for building a power diagram, in which the atoms are added into the system sequentially, one by one. Although very simple, this algorithm is prone to numerical instability. Numerical problems can arise, in particular, when four or more atoms lie approximately in the same plane. (This situation is typical for the macromolecules containing aromatic rings). In this case, the corresponding power cells often have vertices situated at very large distances compared to the dimensions of the molecule. The coordinates of these vertices have low precision, which can cause serious errors by the identifications of the accessible area boundaries on the atom surfaces.

However, as we are interested only in the region near the simulated molecule, there is no need to construct the power diagram in the whole space. The numerical stability can be essentially improved if the construction is limited to some fixed parallelepiped box comprising the molecule. More precisely, we add to the system six auxiliary virtual atoms that correspond to the six faces of the box. They are formally defined in such a way that each individual atom of the molecules, taken alone in their presence, has the power cell that coincides with the given box. This trick improves not only numerical stability, but also the efficiency of the algorithm, as it reduces the total number of the cell vertices.

Another numerically unstable situation occurs when two cell vertices lies within a too short distance from each other as compared with the machine error. One cannot guarantee that, in this case, the topology of the power diagram can be found correctly. As the atoms are added into the system one by one, the easiest way to solve this problem is to decrease the radius of the new coming atom by some value that is small enough to have a negligible effect on the resulting accuracy but, on the other side, is sufficiently large to provide a safe distance between cell vertices.

The third source of the instability is a cell vertex with a near-zero power distance to the four surrounding atoms. That means that the surfaces of all these atoms pass very close to this point, so that their triple intersections cannot be safely resolved. In principle, this situation can be avoided, like before, by a small change of the radius of the new coming atom. However, in our current implementation, we chose another solution, because the construction of the power diagram is performed in a separate module, whereas this problem occurs at later stages of the computation procedure, when the power diagram for the whole molecule is already given. We make use of the fact that the power diagram remains unmodified if all the squares of the radii ρ_i^2 are changed simultaneously by the same value. This change, however, directly affects all power distances, including that of the cell vertices. Thus, if the described situation involves some atom n, then, for the

77

construction of the contours on the surface of this particular atom, we temporally increase the ρ_i^2 values of the neighboring atoms by a small quantity that solves the problem without noticeable effect on the resulting accuracy.

4   Implementation

We implemented our approach in a C++ header-only template library, POWERSASA, intended for computation of molecular surface area, volume and their derivatives. Both float and double precision are provided. As was shown in Ref. 4, our software is essentially more efficient in terms of CPU time, accuracy, and stability in comparison with two other packages, ASC[8] and ALPHASURF[9], currently used for exact computation of the same quantities. More information about POWERSASA is available from the authors on request.

Acknowledgments

References

1. B. Lee and F. M. Richards, *The interpretation of protein structures: Estimation of static accessibility*, J. Mol. Biol. **55**, 379–400, 1971.
2. D. Eisenberg and A. McLachlan, *Solvation energy in protein folding and binding*, Nature **319**, 199–203, 1986.
3. Y. K. Kang, G. Némethy, and H. A. Scheraga, *Free energies of hydration of solute molecules. 1. Improvement of the hydration shell model by exact computations of overlapping volumes*, J. Phys. Chem. **91**, 4105–4109, 1987.
4. K. V. Klenin, F. Tristram, T. Strunk, and W. Wenzel, *Derivatives of molecular surface area and volume: Simple and exact analytical formulas*, J. Comput. Chem. **32**, 2647–2653, 2011.
5. M. L. Connolly, *Analytical molecular surface calculation*, J. Appl. Cryst. **16**, 548–558, 1983.
6. R. Fraczkiewicz and W. Braun, *Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules*, J. Comput. Chem. **19**, 319–333, 1998.
7. H. Edelsbrunner, *The union of balls and its dual shape*, Discrete Comput. Geom. **13**, 415–440, 1995.
8. `http://mendel.imp.ac.at/studies/asc_download.jsp`
9. `http://nook.cs.ucdavis.edu/~koehl/ProShape/download.html`

Prediction of Experimental Phi Values in Protein Folding by Simulation with Knowledge-based Potentials: B Domain of Protein A

Sebastian Kmiecik, Maciej Blaszczyk, and Andrzej Kolinski

Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw
Pasteura 1, 02-093 Warsaw, Poland
E-mail: sekmi@chem.uw.edu.pl

Here we highlight our recent studies which showed that simulation of protein folding process, with the use of knowledge-based potentials and reduced representation of the polypeptide chain, can be a useful method in prediction of experimental Phi values. Apart from providing hints for experimental design, the simulation method (by delivering trajectory of conformations) enables interpretation of the Phi values. The interpretation is a non-trivial task, especially as the obtaining of reliable Phi values is not-trivial either. Here, we present a protein chain mobility profile, from simulation of B domain of protein A, consistent with a detailed Phi value analysis.

1 Introduction

Our recent works demonstrated the usefulness of protein folding simulation with knowledge based potentials for structural characterization of the folding process and denatured state ensemble[1–3]. Such a force-field approximation together with approximations of protein chain representation[4] seem to be a useful tool for qualitative studies of protein folding dynamics of large proteins (not accessible to all-atom Molecular Dynamics - a classical simulation method).

Experimental characterization of protein folding dynamics close to atomic resolution remains a big challenge. The detailed characterization of entire folding process is difficult for small proteins and usually impossible for larger. The only experimental technique sufficiently close to atomic resolution is a protein engineering method called Phi value analysis. The method sometimes leads to misleading results, requires much effort (extensive mutation scanning is needed) and reliable results may be expected only under certain conditions[5].

Giving some hints on the involvement of a particular residues in protein structure formation, Phi value analysis doesn't provide complete information on three-dimensional structure of transient conformers. In the past decade, interpretation of the experimental results leading to structural models[6] became possible thanks to simulation techniques (mainly all-atom Molecular Dynamics) utilizing experimental data and/or knowledge of an experimental (folded) structure.

2 Folding Chain Mobility of B Domain of Protein A and its Correspondence to Phi Values

Recently, while testing the hypothetical mechanism of chaperonin action, we described in detail the folding of B domain of protein A[3]. The results of this de novo simulation
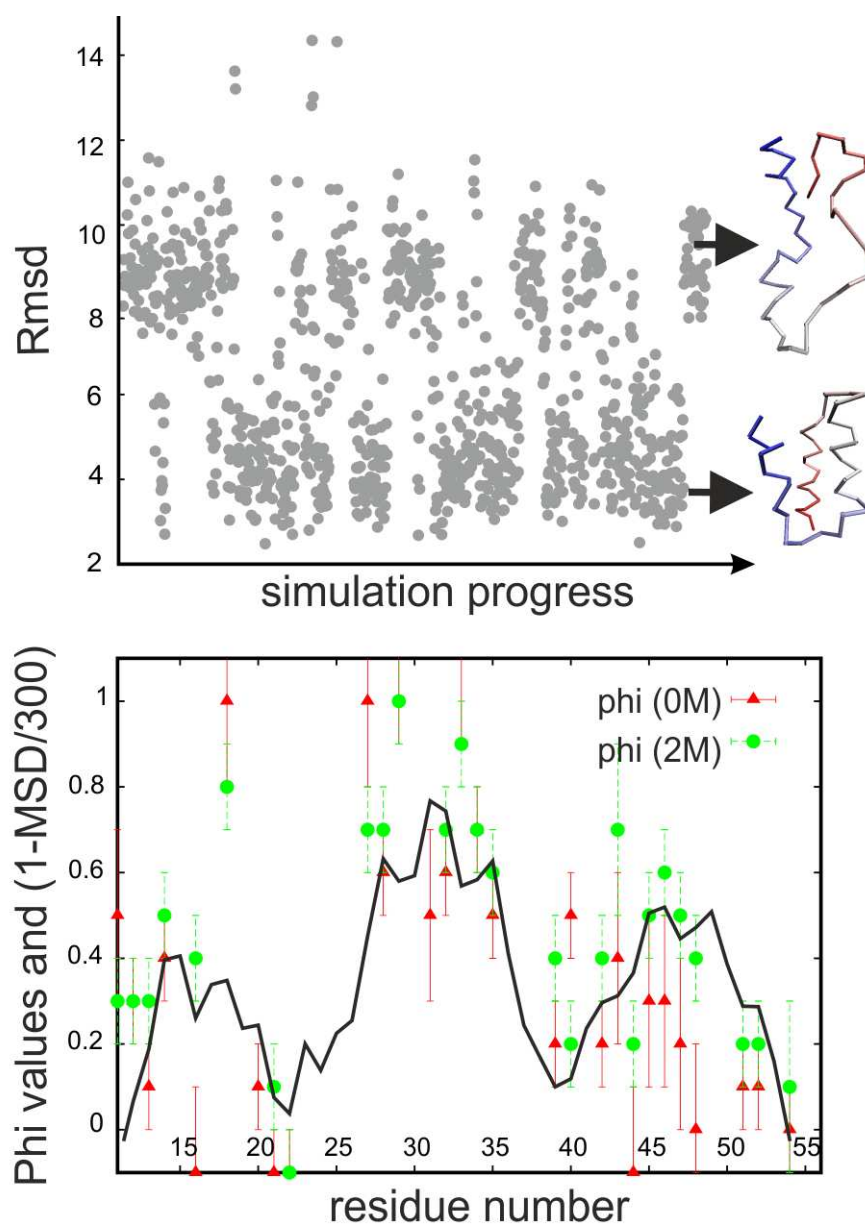
Figure 1. Upper panel: folding trajectory fragment of B domain of protein A (as a function of RMSD from native) together with example conformations from denatured and native like ensemble. Lower panel: profile of the protein chain mobility (marked with the line) superimposed onto Phi values data. The chain mobility is shown as (1-MSD/300), where MSD is a mean square displacement of each residue (average distance a given residue travels along the folding pathway)[3]. Phi values where marked in green and red according to denaturant concentration (0 and 2M GdmCl) together with error bars[7].

(not using any information about the native structure nor any experimental protein-specific data) appeared to be in perfect agreement with the detailed Phi value analysis[7]. This result is quite remarkable taking into consideration: (1) lack of utilizing any experimental data (unlike most of the studies so far), (2) other simulation results - not always agreeing with the experiment[7] and (3) coarse-grained approximations of the CABS model (force-field and protein representation).

Among the paradigm systems of protein folding we studied so far, the transition between the denatured and native like ensemble for B domain of protein A seem to be the easiest to happen along the folding trajectory. Multiple transitions from misfolded (mostly highly expanded) conformations (around 8-11 Angstroms from the native) into native-like (around 5-2 Angstroms from the native) can be observed in a reasonable computational time (in tens of minutes on a single CPU, see the upper panel of the Fig. 1).

As can be seen in the lower panel of the Fig. 1, there is a good correspondence between the protein chain mobility and Phi values – high Phi values associate with low mobility and low Phi values with high mobility along the folding simulation.

High Phi values appeared to be also very well correlated with the most frequently formed ternary interactions along the folding simulation of B domain of protein A (see the Fig. 2C in the Ref. 3). Frequency of contacts maps, derived from the simulation trajectories, enable easy interpretation of the Phi values[3] on the level of particular residue-residue interactions, as we showed also for the other paradigm systems of protein folding[1,2,8].

It is worth noting that the Phi values reflect the degree of structure nativeness around the mutation site in transition state: Phi values close to zero implies that the local structure around the mutation site is relatively unfolded (non-native), if close to 1 implies the similar structure as in the native state. The protein chain mobility measured as MSD (see description of the Fig. 1) reflects local structure involvement in any interactions, not only the native. Thus, the chain mobility profile is more straightforward measure of any structure formation in the particular region and gives the opportunity for studying impact of non-native interactions on observed Phi values.

Acknowledgments

References

1. S. Kmiecik and A. Kolinski, *Characterization of protein-folding pathways by reduced-space modeling*, Proc Natl Acad Sci USA, **104**, no. 30, 12330–5, 2007.
2. S. Kmiecik and A. Kolinski, *Folding pathway of the b1 domain of protein G explored by multiscale modeling*, Biophys J, **94**, no. 3, 726–36, 2008.
3. S. Kmiecik and A. Kolinski, *Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism*, J Am Chem Soc, **133**, no. 26, 10283–9, 2011.
4. A. Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochim Pol, **51**, no. 2, 349–71, 2004.
5. A. R. Fersht and S. Sato, *Phi-value analysis and the nature of protein-folding transition states*, Proc Natl Acad Sci U S A, **101**, no. 21, 7976–81, 2004.

6. R. D. Schaeffer, A. Fersht, and V. Daggett, *Combining experiment and simulation in protein folding: closing the gap for small model systems*, Curr Opin Struct Biol, **18**, no. 1, 4–9, 2008.

7. S. Sato, T. L. Religa, V. Daggett, and A. R. Fersht, *Testing protein-folding simulations by experiment: B domain of protein A*, Proc Natl Acad Sci U S A, **101**, no. 18, 6952–6, 2004.

8. S. Kmiecik, M. Kurcinski, A. Rutkowska, D. Gront, and A. Kolinski, *Denatured proteins and early folding intermediates simulated in a reduced conformational space*, Acta Biochim Pol, **53**, no. 1, 131–44, 2006.

Coarse-grained Protein Modeling: Dynamics, Folding Pathways and Mechanical Unfolding

Andrzej Kolinski, Maciej Blaszczyk, and Sebastian Kmiecik

Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw
Pasteura 1, 02-093 Warsaw, Poland
E-mail: kolinski@chem.uw.edu.pl

Coarse-grained protein modeling tool, CABS, is used in multiscale modeling of protein dynamics. It is demonstrated that the stochastic (Monte Carlo) dynamics, combined with all-atom refinement of the coarse-grained structures follows observed in experiments folding pathways of small proteins. The model is also used in model studies of chaperonin-assisted protein folding. It is shown that Iterative Annealing Mechanism of chaperonin action, where periodic distortion of the polypeptide chains by non-specific hydrophobic interactions can promote rapid folding and leads to a decrease in folding temperature. It is also demonstrated how chaperonin action prevents kinetically trapped conformations and modulates the observed folding mechanisms from nucleation-condensation to a more framework-like. Finally, the CABS model is used in simulations of mechanical unfolding, providing new ways for interpretation of Atomic Force Microscopy experiments.

1   Introduction

Genomic projects provide huge number of protein sequences. Experimental determination of protein structures, while very successful, stays well behind the sequencing – at the moment the number of known structures is about 75000. Knowledge of three dimensional structures of protein and protein-protein and protein-nucleic acids assemblies is crucial for understanding the molecular mechanisms of life, and consequently for the rational drug design, modern biotechnologies, etc. Fortunately, the much cheaper and faster *in silico* structure determination is now possible for a large fraction of known sequences. Comparative modeling is the most accurate and the most advanced method. Obviously, with the increasing number of the high accuracy experimentally determined structures the range of application of comparative modeling will increase. A more straightforward, template-free, structure calculation of new folds is now possible only for relatively small and topologically simple proteins, although there is a steady progress in this area.

By many years theoretical studies of biomacromolecules have been guided by a well-defined paradigm: **from sequence – to structure – to function**. Recently, an extension of this paradigm becomes evident. More and more researchers try to predict and describe the dynamics of biomacromolecules, especially proteins. Professor Harold Scheraga and his coworkers are among the leaders of this trend. It becomes widely recognized that most of protein functions (enzymatic action, recognition, signaling, transporting, etc.) could not be fully understood without a clear reference to their dynamic context. Thus, the current paradigm of structural biology could be rephrased as: **from sequence** – (via dynamics of folding or biomacromolecular assembly pathways) **to structure – through dynamics – to function**.

2 Motivation

Due to the level of complexity, and the timescales involved, computational study of protein systems by means of classical (atomistic level) molecular dynamics (or stochastic dynamics) are now limited to small systems or/and to localized conformational changes. The problem is summarized in Fig. 1, where the timescales of various dynamic processes are compared with the timeframes accessible to various experimental techniques and a typical timeframe of large-scale molecular dynamics simulations. To make the problem of large scale protein dynamics computationally tractable it is necessary to reduce the level of complexity[1–4] and/or to apply more efficient simulation techniques[5]. Coarse-grained models of protein dynamics span a wide range of designs – from mesoscopic continuous models (like Gaussian Network Models – GNM[6]) through united-residue models (like UNRES[7] or CABS[2]) to hybrid models, where some atoms are treated explicitly, while the remaining atoms are combined into united atoms[8,9]. The first class of models (GNM) are at present applicable to moderate changes of molecular geometry, although the size of the tractable systems is essentially unlimited[6]. The GNM-type models are intrinsically structure-based. Conformational fluctuation can be calculated around a starting structure, usually the native-like state. United residue models enable studies of large conformational transitions, while the system's size limits depend on particular designs of the force fields and sampling schemes. The hybrid models bridge the gap between the united-residue and all-atom simulations. Sometimes the structure-based features are introduced into coarse-grained (and all-atom) models via a specific simplification of molecular interactions. For instance, in so called Go-models[10] only the residue-residue interactions seen in the native state define the protein energy landscape. Such approximation seem to work quite well when applied in simulations of mechanical unfolding[11]. On the other hand, the results of numerous applications of Go-type models to protein folding pathways prediction should be treated with extreme cautions – many proteins do fold through non-native intermediates.

In this contribution we describe a couple of new applications (published and unpublished) of CABS model to prediction of protein dynamics. Previously, CABS has been successfully used in test predictions of equilibrium folding pathways of several small proteins[13–15], including the atomic level reconstruction of folding nuclei. Here, we study effects of external forces/restraints on folding/unfolding dynamics and thermodynamics.

3 Coarse-grained CABS Model

CABS is a coarse-grained protein modeling tool applicable to large scale simulations of protein structure (comparative and *de novo* modeling) protein folding dynamics, mechanisms of multimeric assembly and mechanisms of molecular machinery[2]. Acronym CABS stands for the names of united atoms (residues) employed in the reduced representation of protein conformations (see Fig. 2a). The main chain of a polypeptide is represented by the alpha carbon (Ca) trace, which is restricted to an underlying fine-mesh (spacing: 0.61 Å) simple cubic lattice. Allowing some fluctuation of the canonical Ca-Ca distances the number of allowed orientation of Ca-Ca pseudobonds is equal to 800. This safely eliminates any effects of the lattice anisotropy. Additionally, the model backbone contains pseudoatoms located at centers of Ca-Ca pseudobonds. These support a definition of the main-chain hydrogen bonds. The side chains are represented by up to two united atoms,
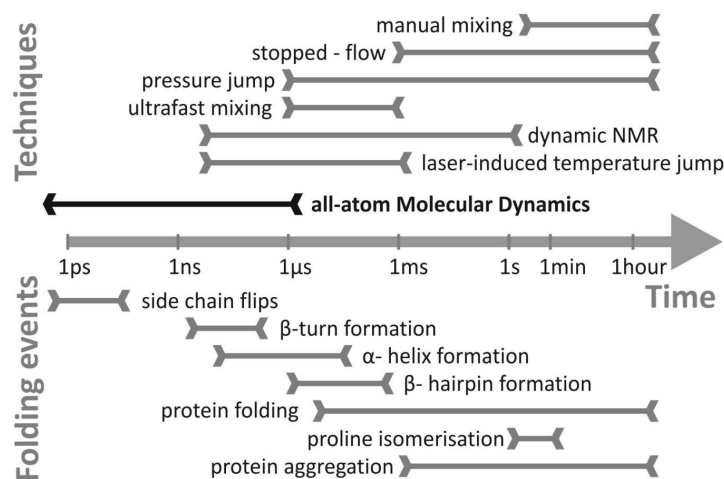
Figure 1. Protein folding dynamics timescale resolutions shown for experimental techniques and all-atom MD (above the axis) and timescales of protein folding events (below the axis)[12].

corresponding to the beta carbons and to geometric centers of the side chains. The side chains are located off-lattice and their positions are defined in the coordinate reference system defined by the Ca-trace (see Fig. 2b), allowing for some rotamer variability.

Motion of CABS polypeptide chains is simulated by means of stochastic dynamics (Monte Carlo Dynamics, MCD) resulting from long series of randomly selected local conformational transitions. These include single Ca moves (see Fig. 2b) accompanied by proper changes of the involved side chains (for clarity the changes of the flanking side chains are not shown in Fig. 2b) and a larger scale transitions, schematically depicted in Fig. 2c. The larger scale moves are attempted less frequently. Lattice representation facilitates very fast computations of conformational transitions leading to speed-up of the model dynamics by about two orders of magnitude in respect to otherwise similar continuous space models. The sampling scheme of CABS depends on specific applications – for structure prediction a multicopy method is recommended (for instance a variant of Replica Exchange Monte Carlo), while for simulations of dynamics a single copy algorithms are easier to interpret (isothermal MCD, or simulated annealing).

Force field of CABS is knowledge-based and consists of statistical potentials derived from structural regularities seen in the known protein structures. The interaction scheme includes: generic short-range conformational propensities mimicking the stiffness of polypeptide chains, sequence-specific short-range (up to four residues along the chain) conformational propensities, a model of highly directional main chain hydrogen bond and the long-range (unbonded) pairwise interactions of the side chains. The side chain interactions are context-specific (depend on mutual orientation of the interacting residues) and thereby account for complex many-body effects (and an averaged implicit solvent effects). Details of the CABS force field design could be found in earlier publications[2] and the numerical data for the potentials could be seen and downloaded from our website (www.biocomp.chem.uw.edu.pl).
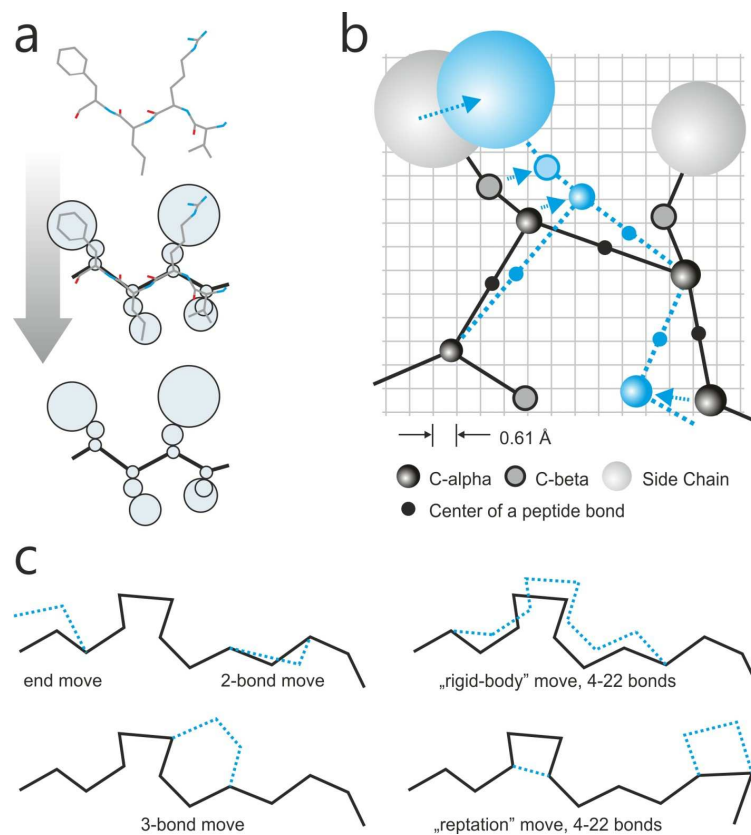
Figure 2. CABS model overview: (a) reduced representation (b) single Cα kink move (c) schematic illustration of larger scale moves of the MCD scheme.

The lattice design of CABS is quite unique. How does it compare with more popular continuous space models? The representation of UNRES[7] is quite similar to that of CABS, except the side chains which in UNRES are single entities in a form of ellipsoids of revolution (two spherical united atoms in CABS). Continuous representation of UNRES enables classical molecular dynamics sampling, although other (sometimes quite sophisticated) sampling schemes were used by the Authors. The main difference is in the designs of the force fields of both models. UNRES force field is physics-based, where parameters of potentials are derived from corresponding atom-level interactions. On contrary, the force field of CABS is fully knowledge-based, attempting to reproduce the very complex patterns of effective interactions in proteins without referring to specific atom-level origins. Both approaches could be comprehensively argued for, depending on a context.

4   Modeling of Chaperonins' Mechanism of Action

Molecular chaperonins are hollow protein chaperones. A single type of a chaperonin (for instance GroEL/GroES complex) can assists in folding of many various proteins. The exact

mechanism of chaperonin action remains largely unknown. That is because difficulties in experimental probing of protein folding mechanisms in such a confined environment. Nevertheless, it is known that for many substrate proteins the interactions with the chaperonin are non-specific and that the periodic changes of hydrophobicity of the chaperonin interior are crucial for its action. Basing on these observations a very simple theoretical model of the chaperonins' mode of action has been proposed[16]. Iterative annealing model (IAM) assumes that the periodic changes of hydrophobicity of the inner walls of a chaperonin complex lead to unfolding of misfolded proteins, consequently enabling them to reach the proper fold. Using the fundamental theoretical concept of IAM we designed a simulation scheme based on CABS. The idea is explained in Fig. 3. The chaperonin is modeled as a spherical object (cylindrical chaperonins were also simulated, without qualitative differences in the results) with thick walls (or a vicinity of the walls) of variable hydrophobicity. In the basic state 9/10 of the simulation time the walls are inert. Periodically (see the right side of Fig. 3) the walls became hydrophobic, attracting the encapsulated protein chain with the strength typical for the hydrophobic interactions within folded proteins (according to the CABS force field).
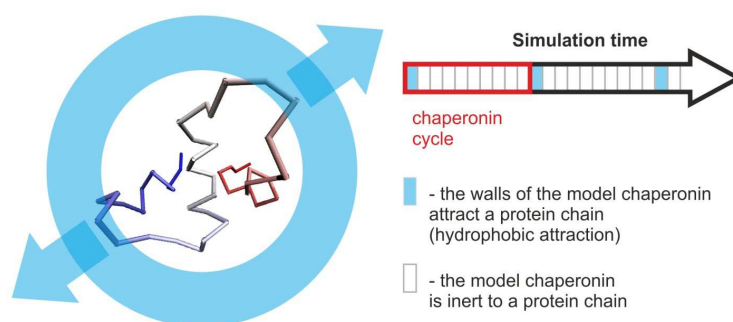

Figure 3. Chaperonin model (see the text).

Chaperonin-assisted folding simulations were performed for two examples: 56 residue B1 domain of protein G and for 46 residue B domain of protein A[16]. The first is an alpha/beta protein and the second a helical protein. These proteins perhaps are not typical chaperonin substrates, but this seems to be irrelevant for the general question about the molecular implications of the IAM model. On the other hand these proteins are paradigms of experimental and theoretical studies of protein folding mechanisms. Since their chaperone-free CABS simulations agree very well with experimental facts, we expect that the observed changes in folding mechanisms within the chaperonin are meaningful. During the simulations we monitored the changes of protein coil/globule volumes, the numbers of native contacts, the secondary structure contents and other structural, dynamic and thermodynamic characteristics. In comparison to the bulk (chaperonin-free) folding simulations the results of simulations of the chaperonin-assisted folding of both proteins could be characterized as follows:

1. Chaperonin significantly (up to two-times) increases the frequency of the folding-unfolding transitions between the near-native structures and the highly denatured structures.

2. Chaperonin decreases the folding temperature (defined as the temperature at which the number of native-denatured transitions is the largest).

3. Chaperonin shifts the folding mechanisms from a sequential, nucleation-and-growth, towards a more prefabricated, framework type assembly. This manifests itself in weakening of typical nucleation sites and a larger volume of intermediates with a significant, native-like, secondary structure content (see Fig. 4).

4. Contrary to the original interpretation of the theoretical IAM model, the simulations show that the chaperonin prevents formation of misfolded conformations, rather than unfolding the already existing misfolds (see Fig. 4).

While very plausible, the specific predictions emerging from these simulations await experimental verification.
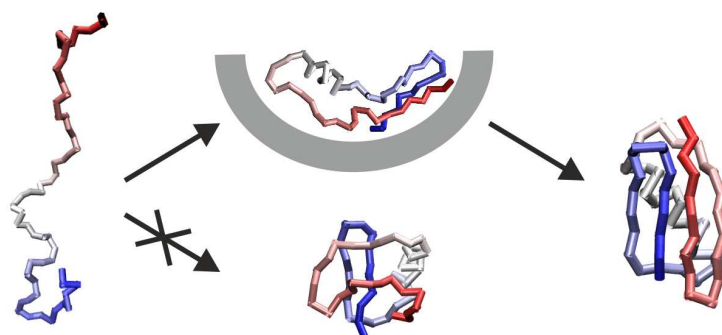

Figure 4. Illustration of the misfolding prevention mechanism within a chaperonin. Pictures from B1 domain of protein G folding simulations[16].

5 *In silico* Atomic Force Microscopy Experiment

The Atomic Force Microscopy is a powerful method for single molecule biophysical experimentations. However, interpretation of the AFM curves requires a model and simulations. Probably, at present the most popular are simulations with Go-type reduced models. As mentioned before, the Go-type approximation, where only native interactions are taken into account, could be non-realistic. Here, we give an example of *in silico* AFM simulation for the CABS model, where the force field do not include the *a priori* knowledge of a specific structure. The simulations correspond to isothermal conditions, well below unfolding temperature. The computational experiment is illustrated in Fig. 5. The molecule (apo-plastocyanin) is attached at one end to the AFM table and at the opposite end to

the molecular cantilever. The cantilever moves apart the table with constant speed and the force exerted on the spring is measured. The resulting force/time (or force/extension) curve exhibits characteristic rapid drops of the force upon unfolding of particular fragments of the structure. Usually, the C-terminal hairpin of the plastocyanin unfolds as the first, immediately followed by unfolding of the N-terminal hairpin (both events are close in time and sometimes the order of the two events changes). The most resistant is the central foldon (a six-member beta barrel) that always unfolds as the last.
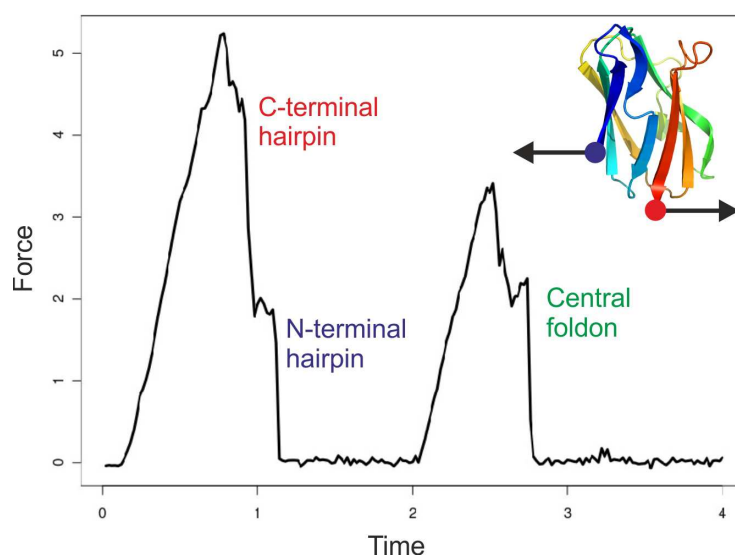
Figure 5. *In silico* AFM experiment with CABS (see the text).

While promising, the CABS assisted interpretation of the AFM experiments requires a precise scaling between the time of the physical AFM experiments and the time of the coarse-grained AFM simulations (which determines a meaningful elongation speed in the simulated experiments). The work along this direction is in progress.

Acknowledgments

References

1. H. A. Scheraga, M. Khalili, and A. Liwo, *Protein-folding dynamics: overview of molecular simulation techniques*, Annu Rev Phys Chem, **58**, 57–83, 2007.

2. A. Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochim Pol, **51**, no. 2, 349–71, 2004.

3. V. Carnevale, C. Micheletti, F. Pontiggia, and R. Potestio, *Bridging the Atomic and Coarse-Grained Descriptions of Collective Motions in Proteins* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 159–178, Springer, New York, 2011.

4. M. Feig, S. Gopal, K. Vadivel, and A. Stumpff-Kane, *Conformational sampling in structure prediction and refinement with atomistic and coarse-grained models* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 85–109, Springer, New York, 2011.

5. U. H. E. Hansmann, *Sampling Protein Energy Landscapes The Quest for Efficient Algorithms* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 209–230, Springer, New York, 2011.

6. T. Haliloglu, I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins*, Physical Review Letters, **79**, no. 16, 3090–3093, 1997.

7. A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Oldziej, K. Wachucik, and H. A. Scheraga, *Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins*, J Phys Chem B, **111**, no. 1, 260–85, 2007.

8. S. M. Gopal, S. Mukherjee, Y. M. Cheng, and M. Feig, *PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy*, Proteins, **78**, no. 5, 1266–81, 2010.

9. A. Irbäck and S. Mohanty, *Effective All-Atom Potentials for Proteins*, pp. 111–126, Springer, New York, 2011.

10. H. Taketomi, Y. Ueda, and N. Go, *Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions*, Int J Pept Protein Res, **7**, no. 6, 445–59, 1975.

11. J. I. Sulkowska and M. Cieplak, *Selection of optimal variants of Go-like models of proteins through studies of stretching*, Biophys J, **95**, no. 7, 3174–91, 2008.

12. S. Kmiecik, M. Jamroz, and A. Kolinski, *Multiscale approach to protein folding dynamics*, in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 281–294, Springer, New York, 2011.

13. S. Kmiecik and A. Kolinski, *Characterization of protein-folding pathways by reduced-space modeling*, Proc Natl Acad Sci USA, **104**, no. 30, 12330–5, 2007.

14. S. Kmiecik and A. Kolinski, *Folding pathway of the B1 domain of protein G explored by multiscale modeling*, Biophys J, **94**, no. 3, 726–36, 2008.

15. S. Kmiecik, M. Kurcinski, A. Rutkowska, D. Gront, and A. Kolinski, *Denatured proteins and early folding intermediates simulated in a reduced conformational space*, Acta Biochim Pol, **53**, no. 1, 131–44, 2006.

16. S. Kmiecik and A. Kolinski, *Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism*, J Am Chem Soc, **133**, no. 26, 10283–9, 2011.

An Optimized Replica Exchange
Molecular Dynamics Approach

Maksim Kouza and Ulrich H. E. Hansmann

Department of Physics, Michigan Technological University, Houghton, MI 49931, USA
E-mail: {mkouza, hansmann}@mtu.edu

We have developed a new extended replica exchange method for studying thermodynamics of
protein folding. It is based on combination of rejection-free moves with the traditional exchange
moves and leads to a more efficient sampling relative to the standard replica exchange method.
Using all-atom model and explicit solvent for simple trp-cage protein we have shown that our
new hybrid approach increases the number of round trips by 25%.

1   Introduction

Computational thermodynamics studies of proteins are very expensive due to long equili-
bration times at low temperatures. To speed up computation of thermodynamics quantities
replica-exchange method has been successfully used[1,2]. The replica exchange is performed
between states with different temperatures allowing to enhance the sampling of low-energy
configurations. However, even for small peptides solvated in explicit solvent one major
problem remains: a huge number of replicas are required to cover the range between the
lowest and largest temperatures. The number n_R of round trips between both temperatures
defines a lower bound on the number of independent configurations sampled at the lowest
temperature. As the round trip times increases as $\sqrt{M}$ with the number of replicas M,
the number of independent sampled configurations decreases accordingly. Hence, protein
simulations with explicit water require not only many replicas but also long simulation
times in order to obtain sufficient statistics at temperatures of interest. In our present work
we test a new approach which allows to reduce computational time by 25% compared to
standard REMD.

2   Methods and Results

In canonical replica exchange[1,3,4] two configurations with energies E_1 and E_2, sitting at
temperatures T_1 and T_2, are exchanged with probability $\exp(\Delta\beta\Delta E)$. Here, we have
introduced the inverse temperature $\beta = 1/k_B T$. These exchange moves generate for each
replica a random walk in temperature that allows the replica to escape local minima. Hence,
sampling becomes more efficient than putting all computing resources in a simulation at
the lowest temperature[1].

In a molecular dynamic simulation, the energy

$$E(x,v) = E_{pot}(x) + E_{kin}(v) \quad \text{with} \quad E_{kin}(v) = \frac{1}{2}\sum_i m_i v_i^2 \qquad (1)$$

is the sum of the potential energy E_{pot}, which depends only on the coordinates x, and the
kinetic energy E_{kin} that is solely a function of the velocities v. Scaling all velocities by a

factor r changes the kinetic energy by:

$$E_{kin}(rv) = r^2 E_{kin}(v) . \tag{2}$$

In standard replica exchange molecular dynamics this is used to scale the velocities after a successful exchange by a factor

$$r_{(1,2)} = \sqrt{T_{(2,1)}/T_{(1,2)}} , \tag{3}$$

that depends on the temperatures T_1 and T_2 of the two replicas that are exchanged. The rescaling of the velocities leads to $v_{(1,2)}^{new} = v_{(2,1)}^{old}$, and therefore $\Delta E_{kin} = 0$. Hence, the probability for an exchange is given by

$$w(1 \leftrightarrow 2) = \exp(\Delta\beta\Delta E_{pot}). \tag{4}$$

In Ref. 5,6 we have proposed a different scaling in the context of microcanonical replica exchange simulations. We scale the velocity in a way that $\Delta E = 0$, i.e. all exchange moves are excepted. Assuming $E_1 < E_2$ and scaling parameters r_1 and r_2 given by

$$r_{1,2} = \sqrt{\frac{E_{2,1} - E_{pot}(\mathbf{x}_{1,2})}{E_{1,2} - E_{pot}(\mathbf{x}_{1,2})}} = \sqrt{\frac{E_{kin}(\mathbf{v}_{(2,1)} \pm \Delta E_{pot}}{E_{kin}(\mathbf{v}_{1,2})}} , \tag{5}$$

one can exchange the two configurations with probability one since :

$$E_1(\mathbf{x}_1, \mathbf{v}_1) = E_{pot}(\mathbf{x}_1) + E_{kin}(\mathbf{v}_1) = E_{pot}(\mathbf{x}_2) + r_2^2 E_{kin}(\mathbf{v}_2) \tag{6}$$

and

$$E_2(\mathbf{x}_2, \mathbf{v}_2) = E_{pot}(\mathbf{x}_2) + E_{kin}(\mathbf{v}_2) = E_{pot}(\mathbf{x}_1) + r_1^2 E_{kin}(\mathbf{v}_1) . \tag{7}$$

The above scaling leads to rejection-free sampling but at the same time also to a lack of importance sampling[7]. We propose the hybrid method where one exchanges most of the times replicas according to Eq. 4, re-scaling the velocities afterwards by Eq. 3, but every 50th exchange attempt is done with the rejection-free moves that follow from a velocity re-scaling according to Eq. 5.
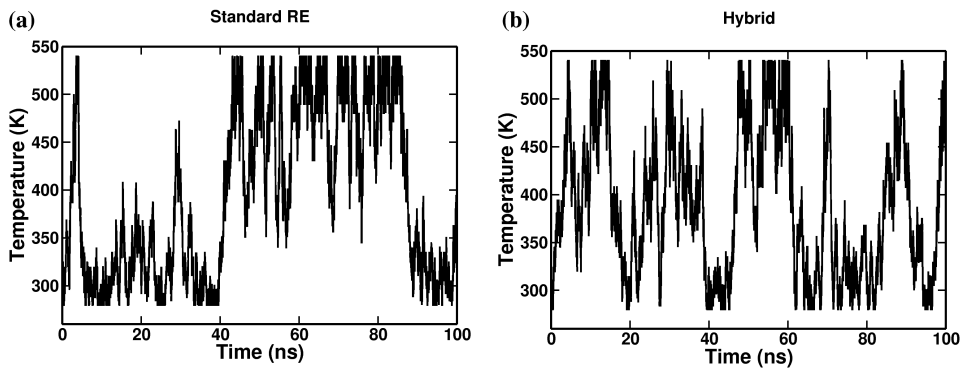

Figure 1. Walk of a specific replica through temperature in (a) a standard canonical replica exchange molecular dynamic simulation, and (b) with our new hybrid approach.

We tested 20-residue trp-cage miniprotein (PDB code 1L2Y) by both methods, standard RE and the hybrid one. The GROMACS program suite[8] was employed using the AMBER94 force-field[9] to model trp-cage protein and the TIP3P water model[10] to describe solvent. The unfolded protein was places in a cubic box with 2645 water molecules. 40 temperatures we used for REMD are taken from Ref.11. More simulation details can be found in works[7,11]

Fig. 1 shows typical replica's trajectories through temperature space for standard (Fig. 1a) and hybrid (Fig. 1b) approaches. The number of round trips between T_1 and T_N, and back, is greater for hybrid method. Over the 100 ns and all replicas we observe 79 and 98 round trips for standard and hybrid methods respectively. This suggest that hybrid approach leads to a more efficient sampling caused by more frequent escaping from local energy minimas.

The distributions of root-mean-square deviation (RMSD) at $T = 280K$ between sampled configurations and the native state one are presented on Fig. 2a. In both cases distributions look very similar and almost all configurations are within 2 Å to the NMR structure. Therefore two methods can find folded structure correctly and hence hybrid approach also leads to correct distribution. This conclusion is also supported by Fig. 2b, where the fraction of folded states is displayed for both approaches. Here, a configuration is defined as folded if its backbone RMSD to the experimentally determined native structure is within 2.2 Å. The measured values obtained by hybrid method are consistent with the those gotten by standard replica exchange molecular dynamics.

3 Conclusion

To improve the efficiency of the replicas's flow through temperature space additional velocity degrees of freedom might be used in replica exchange simulations by using different velocity rescaling equations. The new recently introduced rejection-free moves combined with the traditional ones usually utilized in replica exchange molecular dynamics[1,2] in-
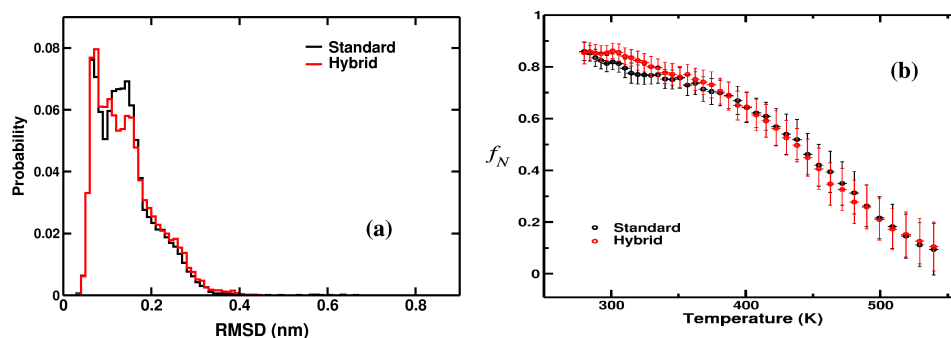
Figure 2. (a) Frequency P of configurations with given root-mean-square deviation as observed at T= 280 K in both standard replica-exchange molecular dynamics, and such relying on mixed the standard and proposed rejection free exchange moves. (b) Fraction f_n of native-like configurations as function of temperatures. Shown are the values obtained from standard replica-exchange molecular dynamics and such as obtained in the hybrid method.

creases by 25% number of round trips. This method might be especially useful in exploring much more rugged energy landscapes of larger proteins.

Acknowledgments

References

1. U.H.E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
2. Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
3. C.J. Geyer and A. Thompson, J. Am. Stat. Ass. **90**, 909 (1995).
4. K. Hukushima and K. Nemoto, J. Phys. Soc. (Jpn.) **65**, 1604 (1996).
5. W. Nadler and U.H.E. Hansmann, Phys. Rev. E. **76**, 057102 (2007).
6. P.Kar, W. Nadler and U.H.E. Hansmann, Phys. Rev. E **80**, 056703 (2009).
7. M. Kouza and U.H.E. Hansmann, J. Chem. Phys **134**, 044124 (2011).
8. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
9. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gouls, D.M. Ferguesson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kolman, J. Am. Chem. Soc. **117**, 5179 (1995).
10. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein, J. Chem. Phys. **79**, 926 (1983).
11. D. Paschek, H. Nymeyer and A.E. Garcia, J. Struct. Biol. **157**, 524 (2007).

Ion Selectivity in VDAC1: A Molecular Dynamics Study of *wild-type* and Mutant Proteins

Eva-Maria Krammer, Fabrice Homblé, and Martine Prévost

Structure and Function of Biological Membranes, Université Libre de Bruxelles
1050 Bruxelles, Belgium
E-mail: {ekrammer, fhomble, mprevost}@ulb.ac.be

The voltage-dependent anion-selective channel (VDAC) forms the major pore for metabolite exchange through the mitochondrial outer membrane. Its open state shows a high conductance and a moderate anion selectivity. It was reported that this selectivity varies upon substitutions of specific charged residues. In this study the mechanism for VDAC selectivity is investigated on the molecular level using molecular and Brownian dynamics simulations of the *wild-type* and mutant VDAC proteins. The analysis of our trajectories shows that two mutants (K252E and E158K) feature a *wild-type*-like ionic distribution, whereas the D30K and K61E variants show altered selectivity. These results agree with the experimental observations. No long-lived protein-ion interactions are identified in the trajectories suggesting that the selectivity does not arise from strong ion binding to specific protein residues. Instead selectivity modulation seems to occur through global changes in the electrostatic pattern of the pore.

1   Introduction

The voltage-gated anion-selective channel (VDAC), the major pore in the outer mitochondrial membrane, is a key regulator of metabolite exchange between mitochondria and cytoplasm. In addition to its pore function, VDAC is also thought to be involved in various cell processes including apoptosis and calcium homeostasis[1]. The physiological significance of VDAC in the mitochondrial metabolism was reported to be strongly correlated to its voltage-dependence. At voltages above 20 mV and below -20 mV, VDAC proteins are known to be (partially) closed and impermeable to metabolites. In contrast, at low voltages the pore is fully opened with a high conductance and a moderate anion selectivity. Experimental measurements[2] and theoretical calculations[3] have indicated that the protein charge distribution is the main factor determining the channel selectivity. Interestingly, this selectivity was reported to vary in yeast VDAC upon substitutions of a few specific charged residues[4].

Recently, the structure of human and mouse VDAC (mVDAC1) was determined at an atomic level[5–7], revealing a β-barrel made of 19 antiparallel strands. These atomic structures allow the use of theoretical methods to address the fundamental principles underlying the ion translocation and selectivity of the pore at a molecular level. In the present study, the effect of mutations on VDAC preference for anions is examined using molecular (MD) and Brownian dynamics (BD) simulations of both, *wild-type* (wt) and several mutant proteins.

2   Methods

Molecular Dynamics simulations. A preexisting equilibrated MD system of mVDAC1 embedded in POPE and in the presence of 0.1M KCl[3] was used as a starting point for

the simulations. After addition of 0.1M KCl and the introduction of point mutations, the system was equilibrated for 12.5 ns using the following sequence: 2.5 ns with all protein atoms fixed, 2.5 ns with all backbone atoms fixed, and 7.5 ns with no restrained atoms. All simulations were carried out using the NAMD program[8] as described elsewhere[3]. For each system a 30 ns trajectory was produced.

Brownian Dynamics simulations. Ten 0.5 μs BD trajectories of the wt and of four mutant proteins were produced under equilibrium conditions and in the presence of 0.2 M KCl using the program gcmcbd and the CHARMM GUI web-based tool[9, 10].

3 Results and Discussion

The ion selectivity mechanism was investigated in yeast VDAC by site-directed mutagenesis replacing individual charged residues with one of the opposite charge[4]. For instance D30K has been reported to feature an increased reversal potential compared to the wt indicating a more marked selectivity whereas K61E mutant showed a decreased reversal potential and a reduced selectivity relative to VDAC wt. Other mutants such as D156K and R252E exhibited a reversal potential similar to the wt. BD simulations[9] performed on the human structure and Poisson-Nernst-Planck calculations[11] carried out on the mouse structure also found that D30K in BD simulations and K61E in both studies alter the channel selectivity.

We generated 30 ns MD and 10x50 μs BD trajectories of the mouse wt protein and D30K, K61E, E158K, and K252E mutants (equivalent to the D30K, K61E, D156K, and R252E variants in yeast) to get insight into the molecular determinants of VDAC selectivity. The time-averaged number of Cl^- and K^+ ions (N_{Cl^-} and N_{K^+}) located inside the pore was computed using the MD and BD trajectories (see Tab. 1). Regardless of the simulation method, the $\frac{N_{Cl^-}}{N_{K^+}}$ ratio is found to be similar in the wt and the K252E mutant. For the E158K mutant this ratio computed using either the MD or the BD trajectory differs. Noticeably the K158 side chain features a high flexibility in the MD trajectory. The lack of protein fluctuations in the BD approach may thus explain this difference and also the disagreement with the experimental observation. The $\frac{N_{Cl^-}}{N_{K^+}}$ ratio (see Tab. 1) increases for the D30K and decreases for the K61E mutant. The ratios between the number of total

	equilibrium MD simulation			equilibrium BD simulation			
subsitution	N_{Cl^-}	N_{K^+}	$\frac{N_{Cl^-}}{N_{K^+}}$	N_{Cl^-}	N_{K^+}	$\frac{N_{Cl^-}}{N_{K^+}}$	$\frac{P_{Cl^-}}{P_{K^+}}$
wt	3.71	1.12	3.33	2.24	0.46	4.88	3.40
D30K	5.01	0.97	5.16	3.73	0.35	10.53	8.23
K61E	3.10	2.55	1.22	1.37	1.15	1.19	0.94
E158K	4.12	1.00	4.12	3.01	0.34	8.86	6.83
K252E	3.69	0.87	4.26	2.00	0.48	4.20	3.22

Table 1. The ratio between the time-averaged number of Cl^- and K^+ ($\frac{N_{Cl^-}}{N_{K^+}}$) inside the pore and the ratio between the number of total crossing events ($\frac{P_{Cl^-}}{P_{K^+}}$) were extracted from the MD and/or the BD trajectories.
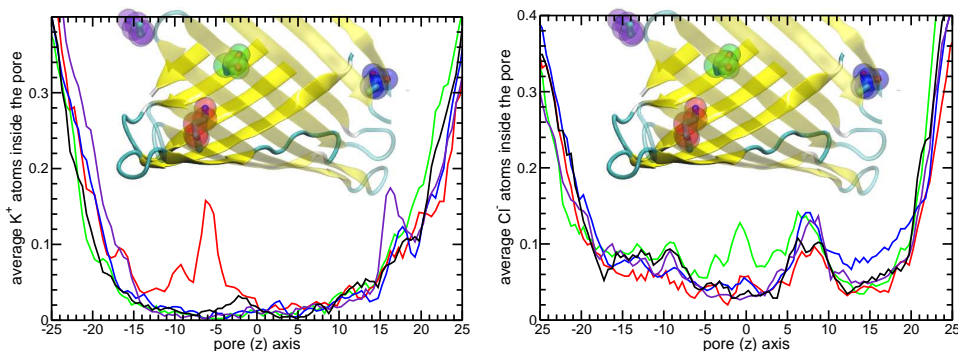
Figure 1. The time-averaged number of K$^+$ (left) and Cl$^-$ (right) ions inside the D30K (green) and K61E (red) mutants differs from the other mutants E158K (blue) and K252E (indigo) and the *wild-type* (black).

crossing events for Cl$^-$ and K$^+$ show the same trends. In the D30K and the K61E mutant MD trajectories an altered ion distribution inside the pore is observed compared to the wt and other mutants (see Fig. 1). In the D30K mutant the time-averaged number of Cl$^-$ ions located inside the pore shows an increased population in a 10 Å wide zone centered on the middle of the pore. In the K61E mutant, the K$^+$ distribution rises in a region located between the entrance and the middle of the pore. These regions were identified to correspond to an energy barrier for the passage of Cl$^-$ and K$^-$ in the wt[3]. Local increase of N$_{K+}$ and N$_{C}l^-$ should thus lead to a decrease in the barriers compared to those in the wt. No long-lived specific interactions between the ions and protein residues could be identified in the MD simulations. We however observe a change in the electrostatic field arising from the distribution of fixed charges inside the pore (see Fig. 2) In the wt anions are attracted from both sides and cations are repelled. This effect is amplified in the D30K mutant whereas it is leveled off in the K61E mutant, in which cations are even attracted on one side.
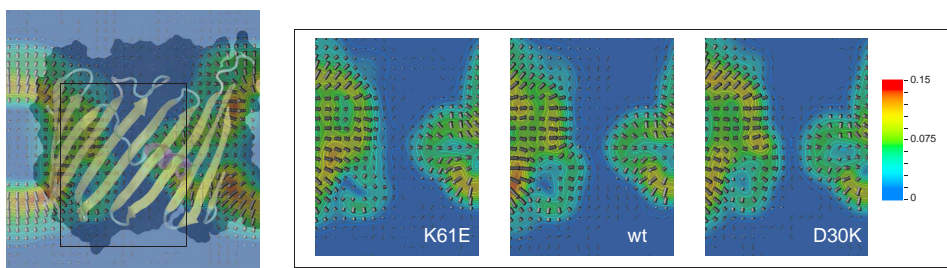

Figure 2. The electrostatic field is altered in the D30K and K61E mutants compared to the *wild-type*.

4 Conclusions

The experimentally observed change in selectivity does not seem to arise from an alteration of specific interactions between the ions and the protein. Instead, our molecular and Brow-

nian dynamics simulations point to the role of the electrostatic pattern of the pore as being the main determinant of the moderate anion selectivity of VDAC. We indeed showed that mutation of specific protein charged residues alters this electrostatic imprint of the pore and thus the protein selectivity. Furthermore the good quantitative agreement between our MD and BD data, except for E158K, suggests that the change in selectivity are not caused by protein fluctuations.

Acknowledgments

References

1. V. Shoshan-Barmatz, V. De Pinto, M. Zweckstetter, Z. Ravivi, N. Keinan, and N. Abel, *VDAC, a Multi-Functional Protein Regulating Cell Life and Death*, Molecular Aspects of Medicinie **31**, 227–285, 2010.

2. E. .B. Zambrowicz, and M. Colombini; *Zero-Current Potentials in a Large Membrane Channel: A Simple Theory Accounts for Complex Behavior*, Biophys. J **65**, 1093–1100, 1993.

3. E.-M. Krammer, F. Homblé, and M. Prévost, *Molecular Determinant of Ion Selectivity in VDAC: a Molecular Dynamics Approach*, *submitted* (2011).

4. E. Blachly-Dyson, S. Peng, M. Colombini, and M. Forte, *Selectivity changes in site-directed mutants of the VDAC ion channel: structural implications*, Science **247**, 1233–1236, 1990.

5. S. Hiller, R. G. Garces, T. J. Malia, V. Y. Orekhov, M. Colombini, and G. Wagner, *Solution structure of the integral human membrane protein VDAC-1 in detergent micelles*, Science **321**, 1206–1210, 2008.

6. M. Bayrhuber, R. Meins, M. Habeck, S. Becker, K. Giller, S. Villinger, C. Vonrhein, C. Griesinger, M. Zweckstetter, and K Zeth, *Structure of the human voltage-dependent anion channel*, Proc. Natl. Acad. Sci. USA **105**, 15370–15375, 2008.

7. R. Ujwal, D. Cascio, J. P. Colletier, S. Faham, J. Zhang, L. Toro, P. Ping, and J. Abramson, *The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating*, Proc. Natl. Acad. Sci. USA **105**, 17742–17747, 2008.

8. J. C. R. Philipps, 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. Comp. Chem. **26**, 1781–1782, 2005.

9. K. .I. Lee, H. Rui,R. .W. Pastor, and W. Im. *Brownian dynamics simulations of ion transport through the VDAC*. Biophys. J. **100**, 611–619, 2011.

10. W. Im, S. Seefeld, and B. Roux. *A Grand Canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels.* Biophys J. **79**, 788-801, 2000.

11. O. P. Choudhary, R. Ujwal, W. Kowallis, R. Coalson, J. Abramson, and M. Grabe. *The electrostatics of VDAC: implications for selectivity and gating.* J. Mol. Biol. **396**, 580–592, 2010.

Chemical Organisation Theory

Peter Kreyssig[1]**, Naoki Matsumaru**[2]**, Florian Centler**[3]**, Pietro Speroni di Fenizio**[4]**, and Peter Dittrich**[1]

[1] Bio Systems Analysis, Department of Mathematics and Computer Science and
Jena Centre for Bioinformatics, Friedrich Schiller University of Jena, 07743 Jena, Germany
E-mail: {peter.kreyssig, peter.dittrich}@uni-jena.de

[2] Gifu University Hospital, Gifu University, Gifu City 501-1193, Japan
E-mail: n_matsu@gifu-u.ac.jp

[3] Department of Environmental Microbiology, UFZ – Helmholtz-Zentrum für Umweltforschung
Permoserstraße 15, 04318 Leipzig, Germany
E-mail: florian.centler@ufz.de

[4] CISUC, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal
E-mail: speroni@dei.uc.pt

The theory of *chemical organisations*[1] is employed as a novel method to analyse and understand biological network models. The method allows to decompose a chemical reaction network into sub-networks that are closed and self-maintaining. Although only stoichiometry is considered to compute organisations, the analysis allows to narrow down the potential dynamic behaviour of the network: organisations represent potential steady state compositions of the system. We review the application to a model of sugar metabolism in *E. coli*[2] and to an HIV dynamics model[3] as well as a theoretical result on fragments of rule based models[4].

1 Introduction – Reaction Networks and Chemical Organisations

By a *reaction network* we mean a pair (M, R) where M is a set and R is a subset of $\mathcal{P}_{mult}(M) \times \mathcal{P}_{mult}(M)$. Here $\mathcal{P}_{mult}(M)$ is the set of multisets over M. We call the elements of M *molecular species* and the elements of R *reactions* resembling the notions of chemistry. By *applying* a reaction $(l, r) \in R$ to a multiset over M we mean replacing the subset l by the subset r.

A being *closed* means that by applying reactions possible to multisets over A we do not get molecules outside A. A being *self-maintaining* means that applying reactions possible at certain rates to a multiset over M does not reduce the number of molecules of any species of A. A subset of M is a *chemical organisation*[1] if it is closed and self-maintaining. An example for the introduced concepts can be found in Fig. 1.

The computation of the chemical organisations can be done using only stoichiometric information. The rate constants and the kinetic law are neglected. But still there is a connection to the *long term behaviour* of the system in the sense that the set of molecules occurring at fixpoints[1], periodic attractors and other limit sets[5] form organisations. This particular property of the chemical organisations is the reason why they can be helpful for structuring bio-chemical reaction networks and for *identifying modules*.
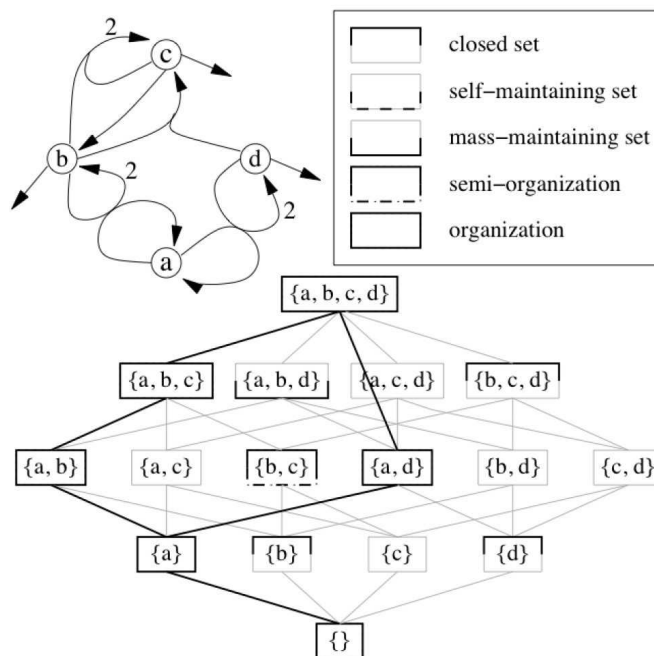
Figure 1. In the upper left corner the network (M, R) with $M = \{a, b, c, d\}$ and $R = \{a+b \rightarrow a+2b, a+d \rightarrow a + 2d, b + c \rightarrow 2c, c \rightarrow b, b + d \rightarrow c, b \rightarrow \emptyset, c \rightarrow \emptyset, d \rightarrow \emptyset\}$ is shown. In the *Hasse diagram* on the bottom several other properties of the subsets are marked. In this example we also see that the organisations form a *lattice*.

2 Analysis of an HIV Dynamics Model

We apply chemical organisation theory to a model[6] describing the interaction of the HI virus with immune system cells[3]. It was developed to explain the efficacy of various drug treatment strategies. Especially it shows, why a drug treatment strategy does not try to remove the virus, but aims at stimulating the immune defence, such that the immune system controls the virus at low but positive quantities. Chemical organisation theory can reveal, even in such relatively small models, a structure (lattice of organisations), which can be used to describe the dynamics of the model and to explain the strategy of a drug treatment from a different perspective.

After the transformation of the ODE model into a reaction network, we can calculate the chemical organisations of the system, see Fig. 2. It shows that the three organisations $\{x\}$, $\{x, y\}$ and $\{x, y, w, z\}$ correspond to "no virus", "immune system destruction" and "virus under control", respectively. It has been observed that although the virus load can be decreased below detection limit, the *virus cannot be fully removed*. Therefore, the actual strategy of a drug therapy is to *move the system into the highest organisation* rather than the lowest, since then the chance is high to move to the middle organisation which represents a destroyed immune system.
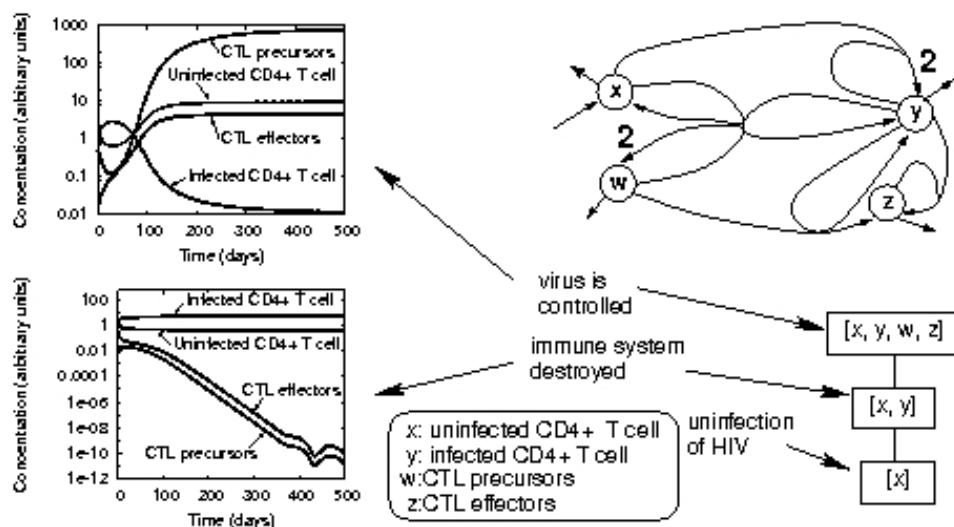
Figure 2. The hierarchy of organisations is shown as a Hasse-diagram (bottom right). Dynamic simulations leading to attractors corresponding to the two upper organisations are shown on the left.

3 Organisations in the Sugar Metabolism of *E. coli*

When looking at organisations of a model of sugar metabolism in *E. coli* including gene expression, signal transduction and enzymatic activities, some are found to coincide with inducible biochemical pathways. This shows that by employing the theory of chemical organisations to intracellular networks it is possible to identify functional units in the network and hence analyse and understand biological network models. The organisations are analysed for several scenarios representing bacterial *growth on different sugar sources* modelled as different inflows, see Fig. 3. The uncovered modules in (**e**) correspond to the *inducible uptake systems* for lactose and glycerol. Therefore the result confirms that *glucose* can be *unconditionally utilised*, while lactose and glycerol can only be utilised after their respective uptake systems have been induced.

4 Fragments and Chemical Organisations

Often the reactions R of a reaction network (M, R) are not explicitly given, but implicitly as *rules* meaning that similar reactions are grouped together. For this we write $(M, \overline{R})$ and refer to it as *rule-based* model. A long term goal is to expand the theory of chemical organisations to this case. The result reviewed here is a first step in this direction.

Fragments[7] form a new *coarse-grained* model out of the rule-based model $(M, \overline{R})$, which is smaller but carries a lot of information about the original set of ODEs. The distinctive feature of fragments is their *soundness*, i.e. assuming mass action kinetics for (M, R) the concentration of a fragment is given as a linear combination of concentrations of molecular species. Therefore fragments help coping with the combinatorial explosion. Since fragments and chemical organisations tackle two different problems concerning
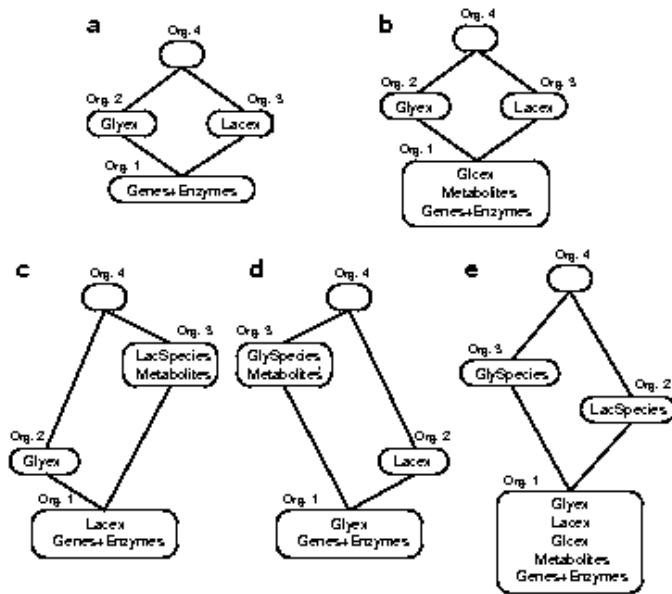
Figure 3. Organisations consist of the species sets contained in their lower organisation(s) plus the species set(s) denoted in their label. (**a**) starvation, no inflow; (**b**) growth on glucose only; (**c**) growth on lactose only; (**d**) growth on glycerol only; (**e**) growth on glucose, lactose, and glycerol.

complexity of reaction networks, it should be *beneficial to combine* them. We showed that in many cases the fragments found in an organisation form an organisation of fragments[4], see Fig. 4. We are planning to deduce some more structural information via this theoretic link, like assertions on quality of coarse graining and easier computation of organisations.
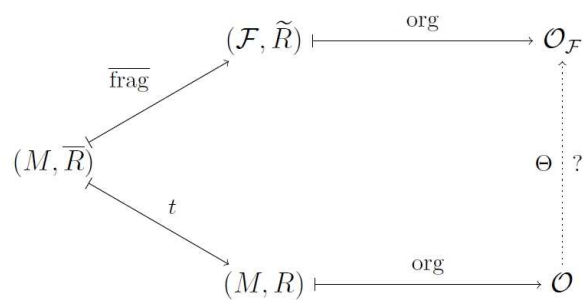

Figure 4. The rule-based $(M, \overline{R})$ and explicit (M, R) reaction network with the set of organisations $\mathcal{O}$ as well as the fragment reaction network and their set of organisations $\mathcal{O}_{\mathcal{F}}$.

Acknowledgments

References

1. P. Dittrich and P. Speroni di Fenizio, *Chemical organisation theory*, Bulletin of Mathematical Biology, **69**, no. 4, 1199–1231, 2007.
2. Florian Centler, Pietro Speroni di Fenizio, Naoki Matsumaru, and Peter Dittrich, "Chemical organizations in the central sugar metabolism of *Escherichia coli*", in: Mathematical Modeling of Biological Systems, Volume I, Modeling and Simulation in Science, Engineering and Technology, pp. 105–119. Birkhuser Boston, 2007.
3. Naoki Matsumaru, Pietro Speroni di Fenizio, Florian Centler, and Peter Dittrich, *A Case Study of Chemical Organization Theory Applied to Virus Dynamics*, in: Systems Biology Workshop at ECAL 2005, Jan T. Kim, (Ed.), Kent, UK, 5-9 September 2005.
4. Peter Kreyssig and Peter Dittrich, *Fragments and Chemical Organisations*, Electronic Notes in Theoretical Computer Science, **272**, 19 – 41, 2011, Proceedings of the 1st International Workshop on Static Analysis and Systems Biology (SASB 2010).
5. S. Peter and P. Dittrich, *On The Relation Between Organizations And Limit Sets In Chemical Reaction Systems*, Advances in Complex Systems (ACS), **14**, no. 01, 77–96, 2011.
6. Dominik Wodarz and Martin A. Nowak, *Specific therapy regimes could lead to long-term immunological control of HIV*, Proceedings of the National Academy of Sciences, **96**, no. 25, 14464–14469, 1999.
7. J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana, *Internal coarse-graining of molecular systems*, Proceedings of the National Academy of Sciences of the United States of America (PNAS), **106**, no. 16, 6453–6458, 2009.

Predicting Protein-protein Interactions with DrugScore^{PPI}: Docking, Scoring, and *in silico* Alanine-scanning

Dennis M. Krüger[1], José Ignacio Garzón[2], Pablo Chacón[2], and Holger Gohlke[1]

[1] Department of Mathematics and Natural Sciences, Heinrich-Heine-University
40225 Düsseldorf, Germany
E-mail: gohlke@uni-duesseldorf.de

[2] Department of Biological Chemical Physics, Rocasolano Physical Chemistry Institute, CSIC
Serrano 119, Madrid 28006, Spain

Protein-protein complexes play key roles in many cellular processes. Therefore, knowledge of the 3D structure of protein-complexes is of fundamental importance. A key goal in protein-protein docking is to identify near-native protein-complex structures using an appropriate scoring function. In this work, we tested our recently developed knowledge-based approach DrugScore^{PPI} (I) for identifying hotspots in protein-protein interfaces, (II) as a scoring function for rescoring a large dataset of pre-generated protein-protein docking decoys, and (III) as an objective function in combination with a fast spherical harmonics-based protein-protein docking tool for predicting 3D structures of protein-protein complexes.

1 Introduction

Protein-protein interactions have important implications in most cellular signaling processes. Therefore, protein interfaces become more and more important as drug targets[1]. Several studies were published addressing the properties of binding sites in protein surface areas and small molecule inhibitors targeting these protein interaction sites[2,3]. From these, it is known that protein-protein complex formation frequently relies on a few interface residues (hotspots) that account for most of the binding free energy. Various protein-protein docking algorithms have been developed during the last years that allow predicting the binding mode of two protein complex partners[4,5]. There are two main aspects in protein-protein docking: (I) Searching for possible docking configurations and (II) assessing each predicted configuration with a scoring function. For the latter, a fast and accurate scoring function is required that identifies a (the) pose that comes close (the closest) to the native structure on the first (on one of the first) scoring rank(s). Despite many advances in the field of protein-protein docking in the past years, the results are still not satisfying: Whereas the generation of several thousands of docking solutions is just a question of computational resources available, the identification of near-native complex structures by a scoring function is still challenging.

In this work, we present a fast and accurate computational approach to predict protein-protein interactions. The approach is based on DrugScore^{PPI}, a knowledge-based scoring function for which pair potentials were derived from 851 complex structures and adapted against 309 experimental alanine scanning results. In our hands, the DrugScore approach has been proven successful already for scoring and predicting protein-ligand and protein-RNA complexes[6,7]. In part, this has been attributed to the implicit, well-balanced consideration of several different types of interactions. Obtaining such a delicate balance is

also considered crucial for successfully predicting protein-protein complexes[8]. To the best of our knowledge, an aspect that has never been addressed is whether these scoring functions are able to predict the change in binding free energy upon single alanine mutations (alanine scanning). Currently, there are only a few computational approaches that address the prediction of hotspots in protein interfaces. Three well-known approaches, FoldX[9], MM/PBSA[10] and the recently published CC/PBSA[11] method, were used for comparison to our QSAR-adapted scoring function DrugScore[PPI] [12].

2 Results and Discussion

2.1 *In silico* Alanine-scanning

DrugScore[PPI] was used for computational alanine-scanning on a dataset of 18 protein-protein complexes with a total of 309 mutations to predict changes in the binding free energy upon mutations in the interface[12]. Computed and experimental values showed a correlation of $r^2 = 0.58$. To improve the predictive power, a QSAR-model was built based on 24 residue-specific atom types. This improves the correlation to $r^2 = 0.73$, with a root mean square deviation of 1.23 kcal/mol. A Leave-One-Out analysis yields a correlation coefficient of $q^2 = 0.64$. For further validation, alanine-scanning with DrugscorePPI was performed on two pharmaceutically important systems, Ras and Raps, which signal to a number of distinct pathways by interacting with diverse downstream effectors. We note that these two systems were not included in the QSAR training set. The results (Tab. 1) demonstrate that DrugscorePPI outperforms other state-of-the-art methods not only with respect to predictive power but also in terms of computational times of 3 seconds per residue on a standard CPU[12].

Method	Ras/RalGDS	Raps/Raf
DrugScore[PPI]	**0.62**	**0.45**
MM-PBSA	0.46	n.a.
FoldX	0.52	0.07
CC/PBSA	0.23	0.22

Table 1. Correlation coefficients of predicted vs. experimental relative binding free energies for the two external test datasets.

Based on these findings, we developed the DrugScore[PPI] webserver[12], accessible at `http://cpclab.uni-duesseldorf.de/dsppi`, which allows identifying hotspot residues in protein-protein interfaces and performing computational alanine scanning of a protein-protein interface within a few minutes.

2.2 Rescoring of Protein-protein Docking Decoys

When used as a scoring function to evaluate decoys of a non-redundant dataset of 54 protein-protein complexes for which "unbound perturbation" solutions have been generated[5], funnel-shaped DrugScore[PPI] score vs. rmsd curves were obtained for the majority
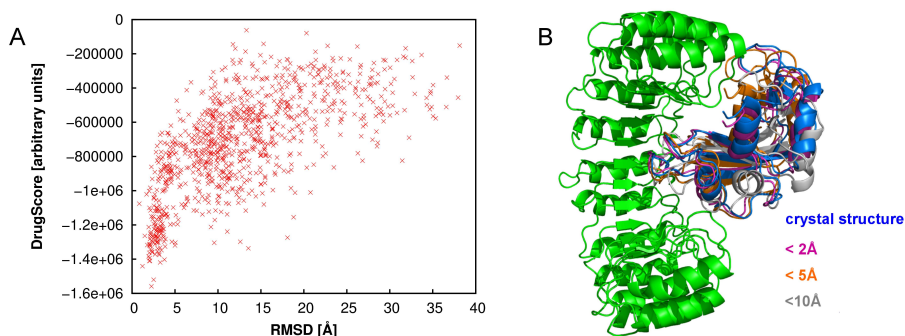
Figure 1. (A) Energy funnel for rescoring docking decoys generated for protein-protein complexes of porcine kallikrein A / bovine pancreatic trypsin inhibitor. (B) Comparison of decoy quality from docking a ribonuclease inhibitor to ribonuclease A. The RMSD is given with respect to the native structure (blue).

of cases (Fig. 1). Accordingly, DrugScore[PPI] was able to rank a near-native solution, i.e. rmsd < 10 Å, in the top ten (five) in 94.1% (64.8%) of the cases. When applied to a dataset of "unbound docking" solutions, DrugScore[PPI] was able to rank a near-native solution in the top ten (five) in 100% (73.3%) of the cases. These results compare favorably with those obtained by Baker et al.[5].

2.3   Protein-protein Docking

Finally, DrugScore[PPI] was applied as an objective function in FRODOCK[4] in order to predict 3D structures of protein-protein complexes. For this, precalculated knowledge-based potential grids by DrugScore[PPI] were used to sample protein-protein configurations and to identify near-native complex configurations (Fig. 1). When applied to a subset of 97 bound-bound test cases of the ZDOCK benchmark 3, convincing results were obtained (docking success rate for complexes in the top ten with rmsd < 10 Å: 69.1%). When comparing the docking results obtained for 76 cases of the ZDOCK benchmark 2 to ZDOCK version 2.3, version 3.0, and the original FRODOCK docking procedure[13,14], our approach outperforms the other methods (Tab. 2).

IRMSD	<=10.0 Å	<=2.5 Å	<= 2.5 Å	< 4 Å
	Top10	Top20	Top100	Top100
DrugScore[PPI a]	**69.1**	**65.6**	**76.0**	**80.2**
DrugScore[PPI b]	**61.8**	**61.8**	**69.7**	**72.4**
ZDOCK2.3	-	18.4	31.6	-
ZDOCK3.0	-	25.0	-	50.0
FRODOCK	-	17.1	30.3	50.0

Table 2. Protein-protein docking results. Success rates in %.

Results achieved for [a]97 cases of ZDOCK benchmark 3 and [b]76 cases of ZDOCK benchmark 2.

Acknowledgments

References

1. Gonzalez Ruiz, D. & Gohlke, H. *Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding*, Curr. Med. Chem. **13**, 2607-2625, 2006.
2. Yura, K. & Hayward, S. *The interwinding nature of protein-protein interfaces and its implication for protein complex formation*, Bioinformatics **25**, 3108-13, 2009.
3. Berg, T. *Small-molecule inhibitors of protein-protein interactions*, Curr Opin Drug Discov Devel **11**, 666-74, 2008.
4. Garzón, J. I., Lopez-Blanco, J. R., Pons, C., Kovacs, J., Abagyan, R., Fernandez-Recio, J. & Chacón, P. *FRODOCK: a new approach for fast rotational protein-protein docking*, Bioinformatics **25**, 2544-51, 2009.
5. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A. & Baker, D. *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*, J Mol Biol **331**, 281-99, 2003.
6. Gohlke, H., Hendlich, M. & Klebe, G. *Knowledge-based scoring function to predict protein-ligand interactions*, J Mol Biol **295**, 337-56, 2000.
7. Pfeffer, P. & Gohlke, H. *DrugScore^{RNA} - knowledge-based scoring function to predict RNA-ligand interactions*, J Chem Inf Model **47**, 1868-76, 2007.
8. Kurcinski, M. & Kolinski, A. *Hierarchical modeling of protein interactions*, J Mol Model **13**, 691-8, 2007.
9. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. *The FoldX web server: an online force field*, Nucleic Acids Res **33**, W382-8, 2005.
10. Fogolari, F., Brigo, A. & Molinari, H. *Protocol for MM/PBSA molecular dynamics simulations of proteins*, Biophys J **85**, 159-66, 2003.
11. Benedix, A., Becker, C. M., de Groot, B. L., Caflisch, A. & Bockmann, R. A. *Predicting free energy changes using structural ensembles*, Nat Methods **6**, 3-4, 2009.
12. Krüger, D.M. & Gohlke, H. *DrugScore^{PPI} webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions*, Nucleic Acids Res **38**, W480, 2010.
13. Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R. & Weng, Z. *Integrating statistical pair potentials into protein complex prediction*, Proteins **69**, 511-520, 2007.
14. Pierce, B. & Weng, Z. *A combination of rescoring and refinement significantly improves protein docking performance*, Proteins **72**, 270-279, 2008.

Survival of the Fattest, the Flattest, or the Fastest?
The Role of Fluctuations in Biological Evolution

Michael Lässig

Institute of Theoretical Physics, University of Cologne, Zülpicher Str. 77, 50937 Köln, Germany
E-mail: lassig@thp.uni-koeln.de

Natural selection is an important factor in biological evolution. This is expressed in the famous Darwinian principle of survival of the fittest. According to this principle, populations should evolve towards peaks of a fitness landscape. However, selection competes with stochastic evolutionary forces, which include mutations and reproductive fluctuations (genetic drift). Moreover, selection itself is often time-dependent and sometimes stochastic: fitness becomes a dynamic seascape rather than a static landscape. Stochastic forces drive populations away from fitness peaks - but where do they end up? In this talk, I discuss fluctuation principles of molecular evolution, which establish links to the statistical physics of entropy and entropy production[1,2]. These principles are applied to the evolution of gene regulation[3,4], of RNA structures, and of the influenza virus.

These applications illustrate the diverse sources and biological consequences of fluctuations in molecular evolution:

Survival of the fattest. Small genomic units are mostly monomorphic, that is, a population can be represented by a point moving in a fitness landscape. At equilibrium, the distribution of population states is shaped by the height and by the fatness of fitness peaks, that is, the number of sequence states with near-optimal fitness. Such equilibria reflect the interplay of selection and genetic drift. An example is the equilibrium distribution of binding energies for regulatory binding sites[3,4].

Survival of the flattest. Quantitative traits depending on a larger number of genomic loci are generically polymorphic, that is, trait and fitness values differ between individuals in a population. Such traits often evolve under significant mutational load. The resulting equilibrium distributions are governed by the height and by the flatness of fitness peaks, that is, the susceptibility of the system to deleterious mutations. An example is micro-RNA processing in plants, a complex function that depends on particular structural elements of the pre-miRNA fold. In this system, the interplay of selection and mutations is shown togenerate genomic modularity, that is, a mechanism for independent evolution of different functions.

Survival of the fastest. Systems under strong adaptive pressure are generically in a non-equilibrium state, and their evolution is shaped by beneficial as well as deleterious mutations[1,2]. In the evolution of the human influenza virus, for example, the viral proteins must adapt to changing host immune challenge. This dynamics is determined by the fastest

strains, which produce beneficial mutations away from host immunity. We show that adaptation leads to interference with protein stability and other viral functions. Adaptation and conservation, i.e., beneficial and deleterious mutations, are tightly coupled in this system, which affects the predictability of influenza evolution.

References

1. V. Mustonen and M. Lässig. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. Trends Genet 25, 111-9, (2009).
2. V. Mustonen and M. Lässig. Fitness flux and ubiquity of adaptive evolution. Proc. Natl. Acad. Sci. 107, 4248-53, (2010).
3. V. Mustonen, J. Kinney, CG. Callan Jr, and M. Lässig. Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. Proc. Natl. Acad. Sci. 105, 12376-81, (2008).
4. M. Lässig. From biophysics to evolutionary genetics: statistical aspects of gene regulation. BMC Bioinformatics 8 Suppl 6, S7, (2007).

Automatic Template-based Model Generation of G-protein Coupled Receptors

Dorota Latek[1] and Slawomir Filipek[2]

[1] International Institute of Molecular and Cell Biology
4 Ks. Trojdena Street, 02-109 Warsaw, Poland
E-mail: dlatek@iimcb.gov.pl

[2] The Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland
E-mail: sfilipek@chem.uw.edu.pl

G-protein coupled receptors are one of the main targets in the pharmaceutical industry. However, only few structures are solved to date and this fact hinders the drug discovery process. In contrast to time-consuming experimental methods of structure determination, homology modeling could be used successfully in many cases, as it was shown in the last GPCR-Dock competition. GPCRs have low degree of sequence similarity, however they share the same sequence motifs and, most importantly, the same seven-helices bundle fold which makes comparative modeling feasible. Here, we present a simple python tool for automatic generation of the alignment and the final model of a receptor. The method is based on profile-profile comparison and well-known tools for template-based modeling such as MUSCLE and Modeller.

1 Introduction

G-protein coupled receptors form the largest family of membrane receptors divided into five main sub-families: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin (the GRAFS classification system)[1]. Among these sub-families, only few Rhodopsin-like receptors have 3D structure solved to date, mostly in the antagonist-bound, inactive state. Structural biology of GPCR confirmed a seven-helices bundle fold, shared by all receptors. Nevertheless, while looking into a more detailed, atomistic picture of GPCR structures, significant differences in side-chain rotameric states of amino acids and slight changes in the backbone of TM helices combined with high structural diversity of extra and intra-cellular loops make homology modeling quite challenging. The typical approach[2] to the structure modeling of GPCRs involves target-template alignment preserving proper assignment of conserved residues and motifs, followed by model building and the last but not the least step of molecular dynamics refinement, most preferably in all-atom lipids and water environment. In this work we focused on the first step, namely the sequence alignment and the model building.

Rhodopsin-like and other families of GPCR were carefully examined by different methods of sequence comparison: traditional phylogonetic analysis[3], HMMs-based comparison[1], multidimensional scaling (MDS)[4] and quite recently also a combined, HHsearch, Needleman-Wunsch-based and motif analyses[5]. These approaches, were, however, mostly used for the classification of GPCRs and not for homology modeling. For the latter task, a multiple sequence alignment approach combined with hydrophobocity and TM region prediction and manual adjustment is more frequently used[2]. Here, we decided to use a well known approach to the target-template alignment which is a profile-profile comparison[6]. The idea of profiles was originally developed by Gribskov[7] and followed by a number of
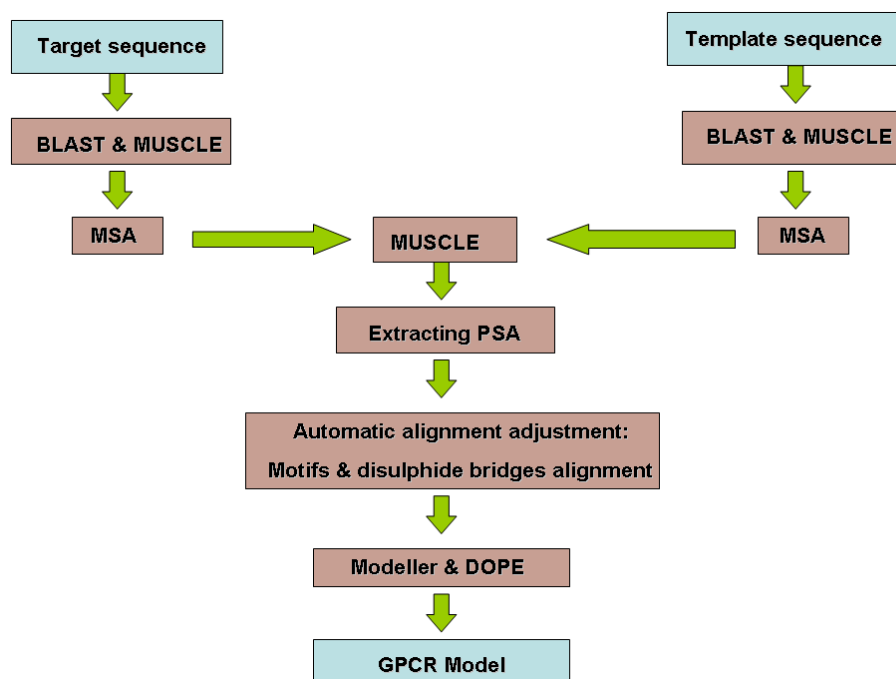
Figure 1. Scheme of the modeling procedure. The crucial step is the target-template profile-profile alignment performed by MUSCLE.

different applications[6,8,9]. Some of the applications use Modeller[10] to reconstruct protein models from the alignments[6] because of its simplicity and high performance. Recently, a new concept was added to the Modeller package - a DOPE scoring function[11], which is a model scoring function based on knowledge-based, statistical potentials.

2   Materials and Methods

The method (see Fig. 1) among several new scripts and programs for processing the sequence alignment utilizes the Biopython libraries[12], a useful, open-source package for biological computations. In the first step of the pipeline the BLAST search[8] is performed for the target and template sequences separately on the non-redundant sequence database. The extracted sequences are aligned by MUSCLE[13] to construct two multiple sequence alignments (MSAs). The MUSCLE program which is used here, is one of the fastest tools for sequence comparison. It utilities fast distance estimation using k-mer counting and progressive alignment followed by refinement using tree-dependent restricted partitioning. Target and template MSAs obtained from the MUSCLE program are merged in profile-profile comparison again by MUSCLE and finally two aligned sequences of the target and the template are extracted. The resulting alignment is automatically checked if all the motifs, conserved residues and the most conserved disulphide bridge between
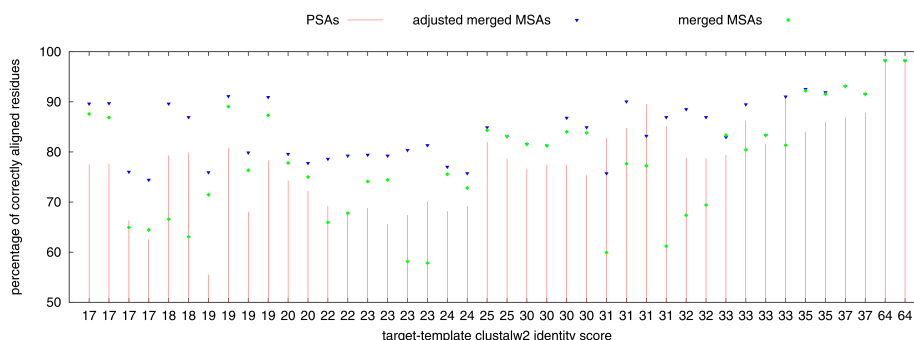
Figure 2. Alignment accuracy vs. target-template sequence similarity. The sequence similarity, shown in x-axis, was computed using Clustalw2 for each pair of the target and template. The alignment accuracy was computed as the ratio of correctly aligned residues divided by the length of the target sequence.

the second extracellular loop (EC2) and TMH3 are properly aligned. Additionally, gaps in the alignment of TM helices which are longer than 1 residue are removed. The alignment is transformed to the PIR format and the DOPE model building with subsequent slow MD-refinement of loops is performed by Modeller. The top 10 DOPE-scored models are selected for comparison with PDB GPCR structures.

3 Results and Discussion

We tested our procedure on all 7 Rhodopsin-like receptors which were solved to date, namely: bovine rhodopsin, beta-1 and beta-2 adrenergic, A2A adenosine, CXCR4 chemokine, dopamine D3 and histamine H1 receptors (PDB id: 1f88, 2vt4, 2rh1, 3eml, 3oe6, 3pbl, 3rze). Since our aim was not to capture the structural difference between the active and inactive state of the receptor we chose only the antagonist-bound structures. At first we computed the alignment accuracy taking the vmd-generated STAMP structural alignment[14] as a reference. In Fig. 2 we compare the accuracy of three alignments: a simple pairwise alignment (PSA) extracted from a single MSA of all 7 GPCRs from PDB with a target sequence, an alignment extracted from two merged MSAs of the target and the template and an adjusted alignment from merged MSAs with all motifs and disulphides properly aligned. A significant improvement of MSAs-involved approach over PSAs is observed in nearly all cases.

From adjusted, MSAs-based alignments with adjustments to maintain the proper alignment of important functional motifs and disulphide bridges, we built 50 protein models by Modeller. The C-alpha RMSD of TM region excluding EC and IC (intra-cellular) loops of the top 10 DOPE-scored models for each target-template pair is presented in Fig. 3. To compare our results with a simple PSAs-based approach we generated 50 models from each target-template PSA and depicted C-alpha RMSD with respect to the native in Fig. 3. Again, a significant improvement of the model quality is observed when using MSAs-based alignments.

Additionally to the above tests, we used our approach in the GPCRDock2010 competition in combination with the Glide docking software[15]. Our model of a small molecule
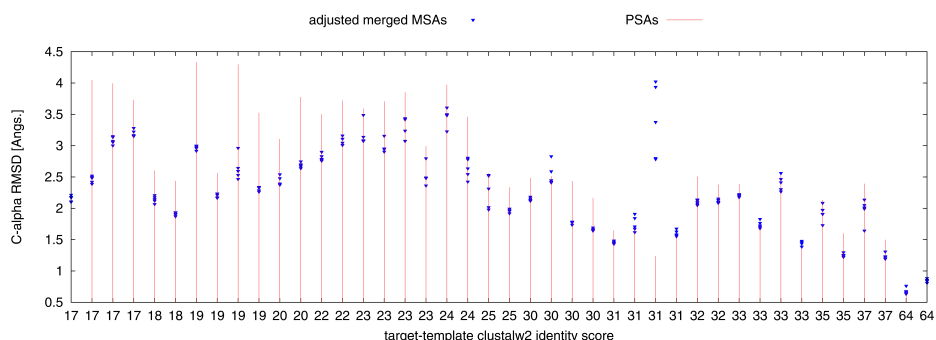
Figure 3. Accuracy of the top 10 DOPE-scored models vs. target-template sequence similarity. C-alpha RMSD was computed using the Bioshell software[16]. To compare, we present also the accuracy of the best single model obtained from each target-template PSA.

eticlopride in a complex with the D3 dopamine receptor was scored as eighth among all sent models[2].

4   Concluding Remarks

The presented method is a first step in GPCRs model building. This stage should be necessarily followed by a MD all-atom refinement, preferably in the explicit membrane environment. The length of loops which can be modeled with the method is limited by the Modeller performance. Since Modeller uses fragment-based approach to loop modeling the loop length should not exceed ca. 10 residues to maintain appropriate performance of the method. The method is freely available for academic purposes and could be downloaded from: `http://www.chem.uw.edu.pl/people/SFilipek/`.

Acknowledgments

References

1. H. B.Schioth, R. Fredriksson, *The GRAFS classification system of G-protein coupled receptors in comparative perspective*, Gen. Comp. Endocrinol **142(1-2)**, 94–101, 2005.
2. I. Kufareva, M. Rueda, V. Katritch, GPCR Dock 2010 participants, C. R. Stevens, R. Abagyan, *Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment*, Structure **19(8)**, 1108–1126, 2011.
3. T. K. Attwood, J. B. Findlay, *Fingerprinting G-protein-coupled receptors*, Protein Eng. **7(2)**, 195–203, 1994.

4. J. Pele, H. Abdi, M. Moreau, D. Thybert, M. Chabbert, *Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors*, PLoS One **6(4)**, e19094, 2011.

5. K. J. V. Nordstrom, M. S. Almen, M. M. Edstam, R. Fredriksson, H. B. Schioth, *Independent HHsearch, Needleman-Wunsch-based and motif analyses reveals the overall hierarchy for most of the G protein-coupled receptor families*, Mol. Biol. Evol., **doi:** 10.1093/molbev/msr061, 2011.

6. L. Rychlewski, L. Jaroszewski, W. Li, A. Godzik, *Comparison of sequence profiles. Strategies for structural predictions using sequence information*, Prot. Sci. **9(2)**, 232–241, 2000.

7. M. Gribskov, D. Eisenberg, *Profile analysis: detection of distantly related proteins*, Proc. Natl. Acad. Sci. USA **84(13)**, 4355–4358, 1987.

8. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*, Nucl. Acids Res. **25(17)**, 3389–3402, 1997.

9. P. Bork, T. Gibson, *Applying motif and profile searches*, Meth. Enzym. **266**, 162–183, 1996.

10. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, *Comparative Protein Structure Modeling Using MODELLER*, Curr. Protoc. Protein Sci., Chapter 2, 2001.

11. M. -y. Shen, A. Sali, *Statistical potential for assessment and prediction of protein structures*, Prot. Sci. **15(11)**, 2507–2524, 2006.

12. P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. L. de Hoon, *Biopython: freely available Python tools for computational molecular biology and bioinformatics*, Bioinformatics **25(11)**, 1422–1423, 2009.

13. R. Edgar, *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*, BMC Bioinfo. **5(1)**, 113, 2004.

14. R. Russell, G. Barton *Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels*, Proteins **14(2)**, 309–323, 1992.

15. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes*, J. Med. Chem. **49(21)**, 6177–6196, 2006.

16. D. Gront, A. Kolinski, *BioShell–a package of tools for structural biology computations*, Bioinformatics **22(5)**, 621–622, 2006.

From Atomistic Simulations to Network Description of Biological Systems

Adam Liwo

Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland
E-mail: adam@chem.univ.gda.pl

Since the advent of high-speed computers, simulations have become an important part of every branch of science and technology, including the biological sciences. The level and extent of detail of simulations depends on the size- and time-scale of the system under study. Individual biomacromolecules and their complexes can be simulated at the atomistic level at comparatively short time scales. Longer time- and size-scales require coarse-grained description in which groups of atoms are merged or, at extreme, a macromolecule is treated as a continuous object. Finally, network description in terms of the influence of the components on each other is used for entire cellular systems. The paper gives an overview of the foundations and applications of these approaches in the context of Harold Scheraga's contribution to the research of biomolecular systems.

1 Introduction

Understanding the functioning of living cells is one of the greatest challenges of contemporary molecular and system biology[1]. The cellular machinery consists of information storage (nucleus or nucleoid), power centers, motors, transcription centers, etc., as well as the building stuff (cytoskeleton), and cell membrane/cell wall. Most of the constituents are composed of biopolymers: nucleic acids, proteins, sugars or assemblies (lipids). The components are related to each other by a network of intermolecular interactions and chemical reactions that give rise to the phenomenon of life.

Even though the elementary interactions belong to the laws of physics, the complexity of a system enables us to use the microscopic approach, namely molecular quantum mechanics, only for very tiny parts of a system. This approach is necessary when studying elementary chemical reactions that are at the root of cellular processes (such as, e.g., protein hydrolysis or reduction of molecular oxygen in the cytochrome system) (Fig. 1). If whole proteins, nucleic acids, enzymes or receptors including the surrounding lipid membranes come into play, it is necessary to resort to empirical force fields[2], in which a molecule is considered at a classical level, and a possibility of a major change of electronic state is ignored. Therefore, the electrons are no longer treated as individual objects and their coordinates are averaged over. On the border line is the QM/MM or Car-Parrinello approach[3], in which a (larger) part of the system is treated at a classical level and a (smaller) essential part in which chemical reactions take place at the QM level.

The atomistic level is still too complex to run large time- and size-scale simulations in real time and to extract important trends from such simulations. Here comes the coarse-graining, a variety of approaches in which a number of atoms are merged into a single interacting sites[4,5]. Coarse-graining can affect just groups of atoms that form an obviously rigid object (such as, e.g., a phenyl group), more flexible objects (e.g., a lysine side chain) or, at the extreme, whole molecules/domains are treated as single interacting sites or a
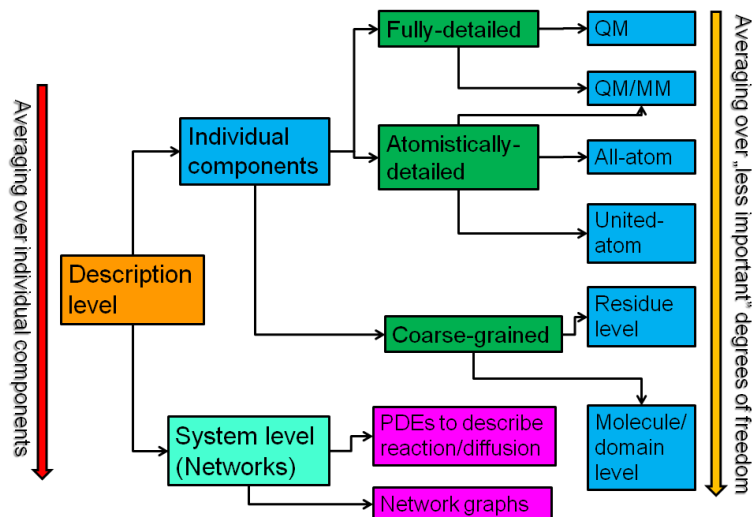
117

Figure 1. Schematic illustration of various views of biomolecular systems.

nucleic acid molecule is treated as an elastic rod (Fig. 1). Explicitly or implicitly, coarse graining means averaging out non-essential degrees of freedom.

All methodologies summarized above belong to component-based approach, in which individual atoms, groups of atoms or molecules are considered. A different view point is offered by a system or network-based approach, in which the whole population of, e.g., a given kind of enzyme is an object and the system is described in terms of "bulk" interactions and dependencies between such collective objects[6] (Fig. 1). The relations can be considered at a detailed chemical level as, e.g., reaction rate constants, or approximated by simple Boolean relations. This view can be compared to that of a general manager of a factory who considers each of its divisions in terms of costs, produce output, efficiency, dependence on each other, rather than taking care of each individual workplace at all times.

In this paper, an outline of the derivation of the above-mentioned approaches to the descriptions of biological systems in terms of averaging over the less important degrees of freedom is presented, and applications of each resolution level are discussed.

2   Atomistically-detailed Approaches

The potential energy surface (PES) of any molecular system can be constructed based on molecular quantum mechanics, according to which the energy is an average of the Hamiltonian operator over system's wavefunction. Within the Born-Oppenheimer approximation, the effective potential energy of a molecular system can be expressed by Eq. 1.

$$U = U\left(\mathbf{R_1}, \mathbf{R_2}, \ldots \mathbf{R_N}\right) = \sum_A \sum_{B<A} \frac{Z_A Z_B}{r_{AB}} + \frac{\left\langle \Psi_{el} | \hat{\mathbf{H}}_{el} | \Psi_{el} \right\rangle}{\langle \Psi_{el} | \Psi_{el} \rangle} \tag{1}$$

with

$$\Psi_{el} = \Psi_{el}\left(\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_N}, \mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_n}, \mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_n}\right) \tag{2}$$

where $\hat{\mathbf{H}}_{el}$ is the electronic part of the Hamiltonian operator (Eq. 4) and Ψ_{el} is the electronic wavefunction that depends on the coordinates of the atomic nuclei $\mathbf{R_1}, \mathbf{R_2}, \ldots \mathbf{R_N}$ and electrons $\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_n}$ present in the system and on the spins of the electrons, $\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_n}$. If only electronic stationary states are considered, Ψ_{el} is a solution of the timeless Schrödinger equation (3) with the electronic Hamiltonian (Eq. 4).

$$\hat{\mathbf{H}}_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{3}$$

$$\hat{\mathbf{H}}_{el} = -\frac{\hbar}{2m_e} \sum_i \Delta_i - \sum_A \sum_i \frac{Z_A}{r_{Ai}} + \sum_i \sum_{j<i} \frac{1}{r_{ij}} \tag{4}$$

where Δ_i denotes the Laplacian in the coordinates of ith electron and m_e electron mass, $\hbar = h/2\pi$, h being the Planck constant.

Eq. 1 provides justification of expressing the PES of a molecular system as a function of nuclear coordinates, which are the essential degrees of freedom here. The electron degrees of freedom are averaged out. Empirical force fields are obtained by approximating Eq. 1 by a neo-classical expression (Eq. 5)[2,7].

$$U = \sum_{bonds} \frac{1}{2} k_i^d \left(d_i - d_i^\circ\right)^2 + \sum_{\substack{bond \\ angles}} \frac{1}{2} k_i^\theta \left(\theta_i - \theta_i^\circ\right)^2 + \sum_{\substack{dihedral \\ angles}} \sum_n \frac{V_i^{(n)}}{2} \left[1 + \cos\left(n\omega_i\right)\right]$$

$$+ \sum_i \sum_{j<i} 332 \frac{q_i q_j}{D r_{ij}} + \epsilon \left[\left(\frac{r_{ij}^\circ}{r_{ij}}\right)^{12} - 2 \left(\frac{r_{ij}^\circ}{r_{ij}}\right)^6\right] \tag{5}$$

where the first two terms account for bond and bond-angle deformation from their unstrained values d° and θ°, respectively, k^d, and k^θ being the force constants, and the third term accounts for torsional interactions, with ω denoting a dihedral angle and $V^{(n)}$ the barrier of the n-fold term. The first term in the double sum expresses Coulombic interactions between the effective charges localized on atoms separated by a distance of r in a medium of relative dielectric constant D, the factor of 332 converting the energy to kcal/mol provided that distances are expressed in angströms and charges in electron-charge unit, the second term in the double sum expresses non-bonded repulsion and dispersion interactions, usually approximated by the Lennard-Jones (12-6) potential, with ϵ and r° denoting the well depth and van der Waals distance, respectively. Nonbonded (electrostatic and Lennard-Jones) interactions are calculated only between atoms that are not bonded or 1,3-nonbonded.

Eq. 5 is approximate, because all terms in it, except the short-range bond-angle and torsional terms are pairwise, while the parent PES given by Eq. 1 cannot be reduced to pairwise terms, even though the potential-energy operators in the electronic Hamiltonian contains pairwise terms exclusively, because of averaging over the degrees of freedom of the electrons. More sophisticated local multibody effects (such as anharmonicity of the bond-stretching potentials, the stretch-bend terms or the bend-torsional terms) are included in class II force fields[7], while long-range "through-space" multibody effects are included in polarizable force fields[7] which are, however, more expensive to use and more difficult to parameterize. Probably the very first force field practically implemented in biomolecular calculations was the Empirical Conformational Energy Program for Peptides and Proteins (ECEPP) developed in Harold Scheraga's laboratory[8].

3   Coarse Graining

The transition from atomistically-detailed to coarse-grained force fields implies averaging out the less important atomic degrees of freedom. A physics-based choice of averaging is the computation of the potential of mean force (PMF) or restricted free energy (RFE) function of the system, as given by Eq. 6[5,9,10].

$$F\left(\mathbf{X}; T\right) \equiv U\left(\mathbf{X}; T\right) = -RT \ln \left\{ \int_{\Omega_{\mathbf{Y}}} \exp\left[-\frac{E\left(\mathbf{X}, \mathbf{Y}\right)}{RT}\right] \right\} dV_{\mathbf{Y}} + C(T) \qquad (6)$$

where $\mathbf{X}$ and $\mathbf{Y}$ denote the primary (coarse-grained) and secondary (averaged out) degrees of freedom, respectively; the two types of the degrees of freedom must be orthogonal to each other, $\mathbf{Y} \perp \mathbf{X}$, T is the absolute temperature, and R is the universal gas constant, $\Omega_{\mathbf{Y}}$ is the space spanned by secondary variables, and $dV_{\mathbf{Y}}$ is the volume element. The additive constant $C(T)$ depends on the reference state selected to compute the PMF[5].

The effective energy function defined by Eq. 6 is directly related to the probability of a coarse-grained configuration of a system. Ensemble averages calculated by using the PMF correspond to those calculated based on the all-atom approach[10], while the expressions for internal energy and heat capacity must contain the first or first and second derivatives of the PMF in temperature to be equivalent to those calculated from the original PES[11].

Eq. 6 can serve a prototype for the derivation of an effective energy function; however, it must be approximated because calculating a multi-dimensional integral over the whole configurational space would cost much more than than atomistically-detailed calculations. This equation is directly implemented in the factor-expansion method[9], which was used to derive the UNRES coarse-grained force field for protein simulations developed in Harold Scheraga's laboratory and in my laboratory[9,12,13] and in the force-matching method developed in Gregory Voth's laboratory[10,14]. Other physics-based approaches, such as the MARTINI force field recently developed by Marrink and coworkers[15] or the very first coarse-grained force field developed for protein simulations by Levitt[16] assume that the effective energy function is a sum of neo-classical terms of the form largely borrowed from Eq. 5; however, such a treatment ignores the multibody terms which are essential when the secondary degrees of freedom are strongly coupled, as occurs in proteins[9]. Consequently, multibody terms are very important to reproduce regular secondary structure of proteins[9].

In the factor-expansion approach, Eq. 6 is expanded into a sum of factors of increasing order (Eq. 7). The factors of order 1 ($f^{(1)}$) comprise the interactions within single sites or between two coarse-grained sites; the latter can be identified with the long-range neo-classical terms, the factors of order 2 contain pairs of interactions such as, e.g., the local interactions within the adjacent amino-acid residues, which give rise to the torsional terms in UNRES[9]. Higher-order terms correspond to more complex correlations such as, e.g., the correlations between backbone-local and backbone-electrostatic interactions in UNRES[9].

$$F(\mathbf{X}) = \sum_i f_i^{(1)}(\mathbf{X}) + \sum_{i<j} f_{ij}^{(2)}(\mathbf{X}) + \sum_{i<j<k} f_{ijk}^{(3)}(\mathbf{X}) + \ldots + \sum_{i_1<2\ldots<i_n} f_{i_1 i_2 \ldots i_n}^{(n)}(\mathbf{X}) \quad (7)$$

Analytical expressions for the first-order factors can be borrowed from Eq. 5 while those for higher-order terms can be derived by expanding a factor into a series of Kubo's generalized cumulants, as in UNRES, to obtain approximate analytical expressions[9]. Each of the factors corresponds only to the sites that are contained in it, out of the context of the whole system. Therefore, the analytical expression can be parameterized by using even high-level quantum mechanical *ab initio* calculations[12]. The factor expansion can also be related to statistical potentials derived from structural database such as, e.g., the CABS force field developed in Andrzej Koliński's laboratory[17] except that it is not possible to separate a given interaction from the context of the entire structure.

The force-matching method, also referred to as the multiscale coarse-graining (MSCG) method[10,14], assumes decomposition of the PMF of Eq. 6 into radial long-range pairwise terms and neo-classical short-range terms. Centers of the interaction sites are located in their centers of masses. The long-range terms are expressed as spline functions in the distance and the force field is parameterized by minimization of the sum of the squares of the differences between the average forces calculated from all-atom MD simulations of a given system and those calculated from the CG model.

Other types of coarse-grained potentials, which are not directly referenced to Eq. 6, comprise arbitrary potentials aimed at studying general features of biopolymers only, structure-based potentials (e.g., the Gō-like potentials), and elastic-network potentials. The reader is referred to recent reviews for details[1,5].

4 Application of All-atom and Coarse-grained Force Fields

The earliest applications of empirical force fields concerned local energy minimization and normal-mode analysis. Search for the most stable structures was initially identified with global minima of the potential energy function[18]. A variety of global-optimization methods has been developed not only for biomacromolecules but also for clusters and crystals, a big part of them in Harold Scheraga's laboratory[8,18]. With the all-atom ECEPP/3 force field and electrostatically-driven Monte Carlo method, the native structure of the 46-residue N-terminal domain of staphylococcal protein A was located as the lowest-energy structure[19]. With the coarse-grained UNRES potential and the conformational space annealing (CSA) method, the structure of phosphate transport system regulator PhoU from *T. maritima* (PDB code: 1SUM; 237 residues)[20] was successfully predicted in 6th Community Wide Experiment on the Critical Assessment of Techniques experiment (CASP6). However, recent findings[11,12,21] demonstrated that conformational entropy is a key player in determining
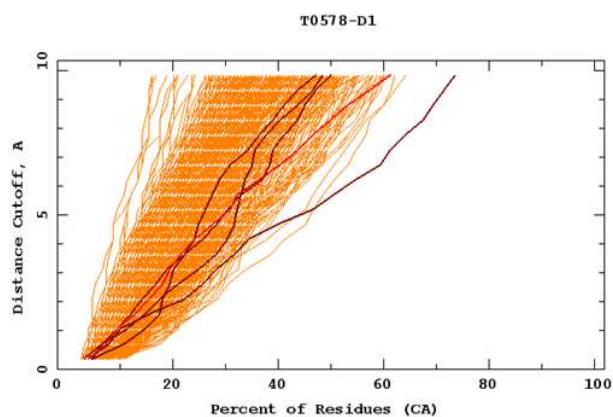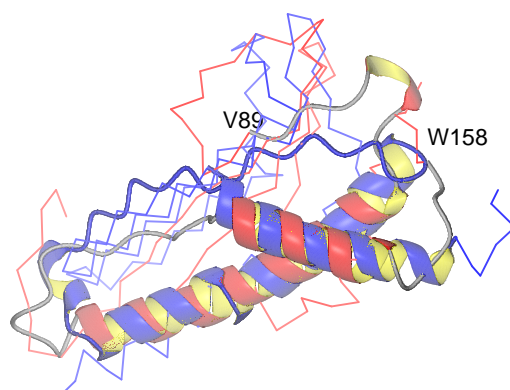
Figure 2. (a) Global Distance Test (GDT) plot of our models of CASP9 target T0578 compared with other groups targets. The rightmost line corresponds to UNRES model 4. (b) Superposition of our model 4 of T0578 (blue) on its experimental structure (red, yellow, and gray). The V89–W158 fragment (ribbon) matches the experimental structure within 5 Å C^α RMSD.

the stability of a structure and the free energy of the entire basin and not the lowest-energy conformation in a basin should be considered. This principle was implemented with the UNRES force field and some success has been achieved in CASP9[13] (Fig. 2).

The most important application of the force fields is studying protein folding. Here molecular dynamics and its extensions are mainly used[1], although appropriately designed Monte Carlo approaches are also applied[17,22]. Recently[23], a dedicated machine ANTON has been constructed, which enables the researchers to carry out all-atom folding simulations with explicit solvent of small proteins (villin headpiece and WW domain) at millisecond scale. Use of the UNRES[12,13,21] or CABS[17,22] force field extends the time- and size-scale to proteins with several hundred residues. Finally, generalized-ensemble sampling with molecular dynamics or Monte Carlo methods as an engine enables us to study the thermodynamics and free-energy landscapes of folding with high confidence[5,12,13].

5 Networks

Although the network view might seem to be an entirely different philosophy, compared to looking at individual components of cellular machinery, averaging provides a bridge between the two approaches. Here, averaging over ensembles of individual components, which can defined as basins in the potential energy surface, is carried out. The essential degrees of freedom remaining are changes of concentrations due to chemical reactions and diffusion. A system is described in terms of a network defining the relations between components, which can be reduced to a graph[6], which can be analyzed by using the apparatus of the graph theory. By comparing the graphs corresponding to different subsystems, differences and similarities in their functioning can be found[6].

6 Conclusions and Outlook

With still increasing computer power, including the machines designed for molecular simulations[23], still greater molecular systems are tractable. At present, it is possible to find most stable structures and to simulate the folding of small (several tens of residues) proteins at atomistic-detail level[8,23]. Coarse-graining enables us to find stable structures and perform folding simulations of proteins with several hundred residues and study conformational changes of even bigger proteins[5,8,13]. Finally, the network-based approaches[6] enable us to get an overview of functioning of the cellular machinery. It is probably a question of time that multiscale approaches will be developed, in which a cellular or subcellular system will be treated simultaneously at all levels, from quantum-mechanical to network, more detailed representation being turned on when important events (such as enzymatic reactions) are expected to occur. The rationale for such a multiscale approach is the averaging The QM/MM approach and simulating protein folding at mixed coarse-grained and atomistically-detailed resolutions, which are already available[3], are the first step to such fully multi-level simulations.

References

1. H. A. Scheraga, M. Khalili, and A. Liwo, *Protein-folding dynamics: Overview of molecular simulation techniques*, Annu. Rev. Phys. Chem., **58**, 57–83, 2007.

2. A. D. Mackerell, Jr., *Empirical force fields for biological macromolecules: overview and issues*, J. Comput. Chem., **25**, 1584–1604, 2004.

3. M. H. Senn and W. Thiel, *QM/MM Methods for Biomolecular Systems*, Angew. Chem.-Int. Ed., **48**, 1198–1229, 2009.

4. G. Voth(editor), *Coarse-Graining of Condensed Phase and Biomolecular Systems*, CRC Press, Taylor & Francis Group, 2008.

5. C. Czaplewski, A. Liwo, M. Makowski, S. Ołdziej, and H. A. Scheraga, "Coarse-grained models of proteins: Theory and applications", in: Multiscale Approaches to Protein Modeling, A. Koliński, (Ed.), chapter 3. Springer, 2010.

6. C. Christensen, J. Thakar, and R. Albert, *Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks*, IET Syst. Biol., **1**, 61–77, 2007.

7. P. Cieplak, F.-Y. Dupradeau, Y. Duan, and J. Wang, *Polarization effects in molecular mechanical force fields*, J. Phys.: Condens. Matter, **21**, 333102, 2009.

8. H. A. Scheraga, *Respice, auspice, prospice*, Annu. Rev. Biophys., **40**, 1–39, 2011.

9. A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga, *Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field*, J. Chem. Phys., **115**, 2323–2347, 2001.

10. S. Izvekov and G. A. Voth, *A multiscale coarse-graining method for biomolecular systems*, J. Phys. Chem. B, **109**, 2469–2473, 2005.

11. A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, and H.A. Scheraga, *Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins*, J. Phys. Chem. B, **111**, 260–285, 2007.

12. A. Liwo, C. Czaplewski, S. Ołdziej, A. V. Rojas, R. Kaźmierkiewicz, M. Makowski, R. K. Murarka, and H. A. Scheraga, "Simulation of protein structure and dynamics with the coarse-grained UNRES force field", in: Coarse-Graining of Condensed Phase and Biomolecular Systems, G. Voth, (Ed.), chapter 8, pp. 107–122. 2008, 2008.

13. A. Liwo, Y. He, and H. A. Scheraga, *Coarse-grained force field: general folding theory*, Phys. Chem. Chem. Phys., DOI: 10.1039/C1CP20752K, 2011, Advance Article.

14. G. S. Ayton, W. G. Noid, and G. A. Voth, *Systematic coarse graining of biomolecular and soft-matter Systems*, MRS Bulletin, **32**, 929–934, 2007.

15. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, *The MARTINI coarse-grained force field: Extension to proteins*, J. Chem. Theor. Comput., **4**, 819–834, 2008.

16. M. Levitt, *A simplified representation of protein conformations for rapid simulation of protein folding.*, J. Mol. Biol., **104**, 59–107, 1976.

17. A. Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochim. Pol., **51**, 349–371, 2004.

18. D. J. Wales and H. A. Scheraga, *Global optimization of clusters, crystals, and biomolecules*, Science, **285**, 1368–1372, 1999.
19. J. A. Vila, D. R. Ripoll, and H. A. Scheraga, *Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures*, Proc. Natl. Acad. Sci. U.S.A., **100**, 14812–14816, 2003.
20. S. Ołdziej, C. Czaplewski, A. Liwo, M. Chinchio, M. Nanias, J.A. Vila, M. Khalili, Y. A. Arnautova, A. Jagielska, M. Makowski, H. D. Schafroth, R. Kaźmierkiewicz, D. R. Ripoll, J. Pillardy, J.A. Saunders, Y.K. Kang, K.D. Gibson, and H.A. Scheraga, *Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests*, Proc. Natl. Acad. Sci. U.S.A., **102**, 7547–7552, 2005.
21. A. Liwo, M. Khalili, and H. A. Scheraga, *Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains*, Proc. Natl. Acad. Sci. U.S.A., **102**, 2362–2367, 2005.
22. S. Kmiecik and A. Kolinski, *Folding pathway of the B1 domain of protein G explored by multiscale modeling*, Biophys. J., **94**, 726–736, 2008.
23. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, *Atomic-Level Characterization of the Structural Dynamics of Proteins*, Science, **341**, 330–346, 2010.

Structurally Non-redundant Protein Sets

Thomas Margraf[1]**, Steve Hoffmann**[2]**, and Andrew E. Torda**[1]

[1] Centre for Bioinformatics, University of Hamburg, 20146 Hamburg, Germany
E-mail: {margraf,torda}@zbh.uni-hamburg.de

[2] Interdisciplinary Center for Bioinformatics and LIFE Research Center for Civilization Diseases
University Leipzig, 04107 Leipzig, Germany
E-mail: steve@bioinf.uni-leipzig.de

We present a method for selecting structurally non-redundant subsets of the PDB. A suffix array based structure search method in conjunction with fast, accurate alignments produces a k-nearest neighbor network. Modularity clustering and other community detection methods decompose the network into clusters from which the most central structures can be selected. The resulting set of protein structures avoids several shortcomings of sequence based sets which are in wide use today. Compared to sequence based methods, we produce significantly fewer and more homogenous clusters.

1 Introduction

Representative lists of proteins are widely used to reduce the sample bias in the PDB. This is essential for all statistical approaches to protein structure analysis and prediction. The various lists in use today either based on classifications of protein structures which do not cover the entire Protein Database, or even more commonly use sequence identity as a filter criterion. Hobohm et. al.[1] state that representative pdb sets provide maximum coverage with minimum redundancy. While their selection is based on sequence similarity, they suggest that: "If the goal is to have a set of structurally unique proteins, then explicit structural superposition should be used, rather than sequence alignment"[1]. The reasons are fairly obvious: similar sequences imply similar structures, but dissimilar sequences do not necessarily imply dissimilar structures. Hence, sequence based lists always contain proteins with similar structures. Furthermore, examples like the Paracelsus challenge show that even the first statement doesn't always hold. Another problem is the case where the same protein has been crystallized in different conformations. Although identical in sequence, the structures may be distinct enough to include both conformations in the representative set. Our method seeks to avoid these problems.

2 Methods

Our method uses an index based structure search method[3] to determine the most closely related structures for each protein. Our structure alignment method relies on a classification of fragments into about 300 classes. Given a fragment (length 6) from a protein, one can calculate the probability of being in each class[7]. For this work, we have a discretized the vectors into an alphabet of about 1000 characters. This loses information, but allows one to use an enhanced suffix array for very fast lookups. After the PDB has been indexed, one can perform lookups in linear time with respect to query length. The k best hits can

then be accurately re-aligned using the original structural data to yield better scores for ranking. Since k is constant and the average size of protein chains is roughly constant as well, every such query can be performed in constant time (about 6 seconds on current hardware, $k = 20$). Therefore, we can use this method to compute the k most similar structures for every chain in the database in time linear to the size of that database. From this, we construct a k-nearest-neighbor graph with nodes for every structure and edges for the similarity relationships. The edge weights can be arbitrary similarity or distance scores. This graph can be decomposed by removing edges above a certain weight threshold and community detection algorithms such as modularity clustering[5]. Modularity clustering optimizes the modularity of a graph. That means it partitions a graph so that there are fewer edges between clusters than one would expect by chance, and more edges within a cluster than in a random graph. The time complexity of this algorithm is $O(n^2 log\ n)$. Modularity clustering does not require a weighted graph, but since not every structure has 20 similar chains in the PDB, computing edge weights and applying a threshold facilitates the graph decomposition. Furthermore, setting a distance cutoff is useful for the interpretation of the resulting clusters.

Picking the highest quality representatives, is a challenge as well. Our approach is to filter and sort the structures according to criteria such as missing coordinates, resolution and sequence length.

Our second method implements the "Select until done" method from Hobohm et. al.[1] using the SALAMI structure search tool[2]. The feasibility of this method depends on the distribution of the PDB entries in structure space. If most structures form tight clusters, the algorithm will converge rather quickly. However, if the distribution is more even, then the complexity approaches that of all vs. all comparisons.

3   Results and Discussion

As a proof of principle, we chose a non-redundant set of 25101 protein chains with pairwise sequence identities less than 90%. The complexity of our algorithms easily allows the clustering of the entire protein databank, but in order to speed up the analysis of the clusters a smaller dataset was required. The edge weights of the similarity graph are RMSD values.

The number of clusters (simply connected components) increases monotonically as we lower the RMSD threshold. The number of biconnected components however drops for cutoff values smaller than 3Å. This can be explained by the fact that a biconnected component needs to stay connected after one edge is removed. This implies that biconnected components contain at least 3 nodes. As the size of the connected components decreases, fewer clusters satisfy that definition. Beyond 10 Å, the numbers of clusters and biconnected components are very similar. This shows that there are very few clusters with 1 or 2 members in this region, and the clusters are still well connected. Such RMSD values might seem too large to be meaningful, but one needs to bear in mind that SALAMI does not perform any postprocessing or refinement of the alignment. Thus SALAMI alignments are much larger and the RMSD values are not directly comparable to those of other structure alignment tools.

The clusters50.txt file distributed with the PDB[6] clusters protein chains with a threshold of 50% sequence identity. This file defines 19599 clusters, a rather small reduction compared to the 25101 clusters at 90% identity. Even at such a conservative threshold,
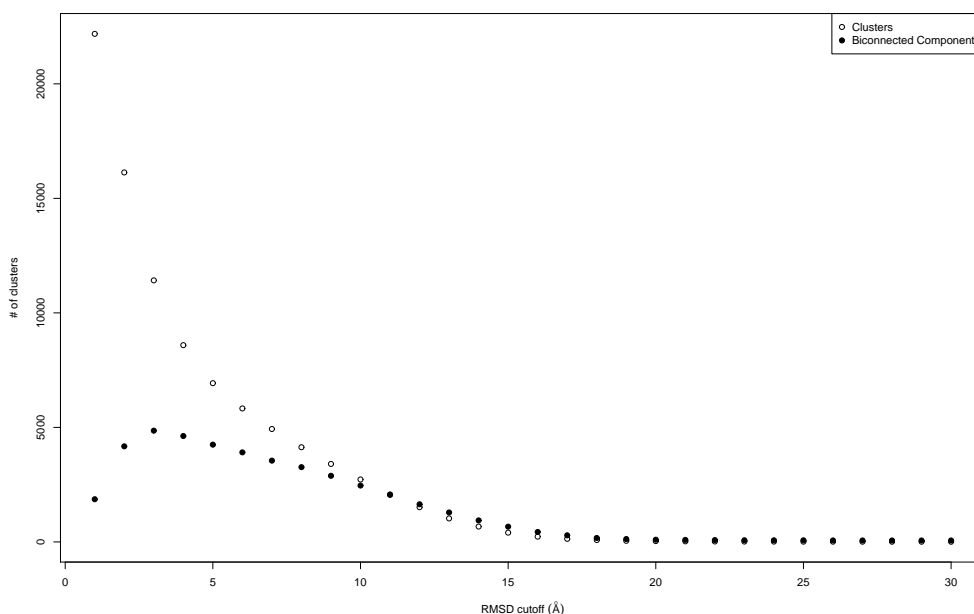
Figure 1. Number of clusters and biconnected components for different RMSD cutoffs of protein structures with less than 90% sequence identity.

25% of alignments within clusters have RMSDs larger than 2.48Å. Our method on the other hand only produces 13636 clusters at a threshold of 2.48Å. This result shows that our method produces significantly fewer clusters which are structurally more homogenous.

The modularity clustering shows the optimal solution for 90 classes. This would correspond to an RMSD cutoff between 17 and 18 Å. It remains to be seen if this really is a natural division of the protein structure space, or simply an artifact due to the fact that our modularity clustering implementation does not yet consider edge weights.

4   Conclusion

In their review of sequence based redundancy reduction methods, Sikic and Carugo[4] found huge overlaps between sets constructed with different methods. Therefore, we do not expect dramatic differences between the structure based lists using the same scoring function. They are however distinct from sequence based lists. It remains to be seen if the expensive alignment method converges in a reasonable amount of time. Furthermore, it will be interesting to see if it produces superior results which justify the huge amount of CPU time. Depending on the desired application, different scores may be used for the thresholds. For example for detecting different conformations within sequence clusters, the RMSD is very well suited. However, its size dependence makes it difficult to set a good threshold for inter-cluster comparisons. Length adjusted alignment scores are much more robust in that respect.

There is a lot of work still to be done. The most important tasks are evaluation of different structural similarity measures, extending the modularity clustering step to take edge weights into account and running the calculations on the whole PDB.

References

1. U. Hobohm, M. Scharf, R. Schneider, and C. Sander. *Selection of representative protein data sets*, Prot. Sci. **1**, 409–417, 1992.
2. T. Margraf, G. Schenk, and A. E. Torda, *The SALAMI protein structure search server*, Nucl. Acids Res. **37**, W480–W484, 2009.
3. S. Hoffmann, *Using index based techniques in protein structure comparison*, University of Hamburg, 2007.
4. K. Sikic, and O. Carugo. *Protein sequence redundancy reduction: comparison of various methods*, Bioinformation **5**, 234–239, 2010.
5. M. E. J. Newman. *Modularity and community structure in networks*, PNAS **103**, 8577–8582, 2006.
6. W. Li and A. Godzik. *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences*, Bioinformatics **22**, 1658–1659, 2006.
7. G. Schenk and T. Margraf and A. E.Torda. *Protein sequence and structure alignments within one framework*, Algorithms for Molecular Biology **3**, 2008.

Continuous-space Sequence Optimisation for RNA Secondary Structures

Marco C. Matthies[1], **Stefan Bienert**[1], **Kristina Gorkotte-Szameit**[2], **Cindy Meyer**[2], **Ulrich Hahn**[2], **and Andrew E. Torda**[1]

[1] Centre for Bioinformatics
[2] Institute for Biochemistry and Molecular Biology
University of Hamburg, 20146 Hamburg, Germany
E-mail: {matthies, bienert, torda}@zbh.uni-hamburg.de
E-mail: {gorkotte, cindy.meyer, uli.hahn}@chemie.uni-hamburg.de

Finding an RNA sequence that will form a chosen secondary structure is usually regarded as an optimisation problem in a discrete sequence space, but it can be useful to let the search take place in a continuous space instead. A differentiable scoring function defined on this continuous sequence space allows the use of optimisation algorithms such as steepest descent or dynamical simulated annealing that rely on gradient information. The scoring function contains a term that models compatibility with the target structure according to the RNA nearest-neighbour energy model, a mean-field negative design term to disfavour unwanted structure formation, constraint terms, and a time-dependent term to force convergence to a solution in the discrete space of physically realisable molecules. We have used the algorithm to design sequences for a three-arm junction and a tRNA-like secondary structure and tested the molecules in the laboratory. The chemical structure probing method SHAPE (Selective 2'-hydroxyl acylation analysed by primer extension) yielded results consistent with the target secondary structure.

1   Introduction

RNA is a molecule with many possible applications ranging from biology to nanotechnology, and it is therefore an interesting problem to find sequences that will fold into a given target structure. In order to computationally test a designed sequence before committing oneself to the laboratory, one must have a reliable structure prediction method. Although there has been significant progress, three-dimensional structure prediction for RNA molecules remains a more speculative undertaking. For RNA secondary structure prediction however, a wide variety of somewhat successful methods is available.

The most popular of these methods are based on the nearest-neighbour energy model, which begins with a decomposition of a given secondary structure into its constituent substructures such as hairpin loops, bulge and interior loops, and multiway branching loops. The free energy of folding an RNA sequence s from the unfolded state into a given secondary structure ω is then approximated as the sum of loop contributions:

$$\Delta G_{\text{fold}}(s,\omega) = \sum_{L\, \in\, \text{Loops}(\omega)} \Delta G_L(s) \tag{1}$$

The advantage of this rather empirical approach to free energy calculation is that, even though the number of possible secondary structures grows exponentially with sequence length, the minimum free energy structure, partition function Z, and base pairing probabilities p_{ij} can be calculated for all unpseudoknotted secondary structures in cubic time

via dynamic programming[1]. For a test set of biological sequences smaller than 700 nucleotides, 73% of known base pairs were correctly predicted by this method[2].

A designed sequence can be evaluated according to three different criteria of increasing rigour: the predicted minimum free energy structure should be the target structure, the ensemble-weighted base-pair distance to the target should be as small as possible, and finally the most stringent criterion is that the probability of the target structure approaches unity.

We can calculate the ensemble-weighted base pair distance $d_s(\omega^\star)$, i.e. the expected number of base pairs that are different to the target structure $\omega^\star$, with the help of the base pair probabilities p_{ij} by

$$d_s(\omega^\star) = \sum_{(i,j),i<j} p_{ij}(1 - \omega_{ij}^\star) + (1 - p_{ij})\omega_{ij}^\star \tag{2}$$

where $\omega_{ij}^\star = 1$ if i and j are base-paired in $\omega^\star$, and $\omega_{ij}^\star = 0$ otherwise. The probability of the target structure $\mathcal{P}_s(\omega^\star)$ can be calculated from the partition function as

$$\mathcal{P}_s(\omega^\star) = Z^{-1} e^{-\Delta G_{\text{fold}}(s,\omega^\star)/RT} \ . \tag{3}$$

2 Methods

Most existing RNA sequence design algorithms approach the task as an optimisation problem in a discrete sequence space. The simplest approach is to iteratively perform local optimisation in this space and then, after each set of modifications, evaluate the designed sequence. This can be repeated as often as necessary to achieve the desired quality.

This approach becomes rather slow for large target structures due to the cubic runtime of structure prediction, and therefore more sophisticated algorithms such as RNAinverse[1], RNA-SSD[3], INFO-RNA[4], and NUPACK[5] use a hierarchical approach to sequence design. The target secondary structure is decomposed into a tree of structural elements (which is always possible for unpseudoknotted structures) and sequences are designed for each substructure at each level of the tree separately. The designed sequences are then merged and taken to be the starting point for sequence optimisation at the next higher level.

Although these methods are significantly more efficient, they nonetheless depend on a hierarchical structure decomposition and therefore seem difficult to adapt to more sophisticated structural representations or models of folding that incorporate kinetics. Another difficulty stems from the fact that for many target structures, the ensemble-weighted base pair distance and the probability of the target structure are global properties of the sequence.

We chose to use a continuous or alchemical representation of an RNA molecule, in which all four base types are present with different probabilities at each position of the sequence and we then optimise the base probability distributions.

Our scoring function defined on these base probability distributions attempts to capture the two most important aspects of sequence design: affinity (positive design) and specificity (negative design) to the target structure. As the scoring function is differentiable with respect to the composition at each position of the sequence, we can use gradient-directed optimisation methods. This allows us to change many positions in the sequence at once in a correlated fashion. Our use of alchemical hybrid residues has some similarities to the λ-dynamics[6] method for free energy calculations and fast ligand-ranking as well as

the ADAPT multicopy molecular mechanics approach of Lafontaine & Lavery[7] to study and optimise DNA sequences for protein binding and preference for structural transitions.

During the final stages of an optimisation run, we add a purification term V_{pur} to our scoring function that favours physically realisable sequences, i.e. the composition at each position should be unity for one base type and zero for all others. A soft constraint term V_{constr} restrains the probabilities to plausible values.

The scoring function has the overall form

$$V(s, \omega^\star, t) = V_{fold}(s, \omega^\star) + V_{nd}(s, \omega^\star) + V_{pur}(s, t) + V_{constr}(s) \qquad (4)$$

The positive design term V_{fold} models affinity to the target structure as a sum of expected loop contributions given the probability distribution $\mathcal{P}(s_L)$ over the sequence in the loop:

$$V_{fold}(s, \omega^\star) = \sum_{L \in Loop(\omega)} \sum_{s_L} \mathcal{P}(s_L) \Delta G_L(s_L) \qquad (5)$$

The negative design term V_{nd} tries to penalise affinity to unwanted structures. We approximate the influence of other possible structures by letting each position in the sequence interact with a mean-field of all other positions it does not interact with in the target structure but could possibly interact with in theory:

$$V_{nd}(s, \omega^\star) = \frac{1}{N} \sum_{(i,j) \in \overline{\omega^\star}} \sum_{k,l} s_{i_k} s_{j_l} E_{nd}(k, l) \qquad (6)$$

Here $\overline{\omega^\star}$ is the set of all possible unwanted interactions, s_{i_k} is the probability of base type k at position i in the sequence, and E_{nd} is a simplified interaction energy between base types without nearest-neighbour effects.

This scoring function can be regarded as a potential energy function, from which equations of motion in sequence space can be derived and numerically integrated. By coupling the "sequence temperature" to a thermostat one can perform dynamical simulated annealing, starting at a high temperature and gradually decreasing it until one converges to a local minimum.

3 Results

We have computationally tested our design method on many target structures. Here we report on two experiments where we tested designed molecules in the laboratory. Sequences for two secondary structures were designed: a three-arm junction and a tRNA-like four-arm junction, both shown in Fig. 1. The sequences were chosen by designing 1000 sequences for each target, filtering those for which the predicted probability of the target structure was below 90%. From the remaining sequences, the one with the lowest GC-content was chosen. The chemical structural probing method SHAPE[8] (Selective 2'-hydroxyl acylation analysed by primer extension) was used to test designed sequences. In SHAPE, an acylating agent preferentially reacts with the 2'-hydroxyl group of unpaired nucleotides. A subsequent reverse transcription with labelled primers will terminate at these acylated nucleotides, allowing one to determine the reactivity at each position in the molecule via gel electrophoresis. We show the SHAPE reactivity at each base superimposed over the target secondary structure in Fig. 1 for illustrative purposes, but we wish to emphasise that the
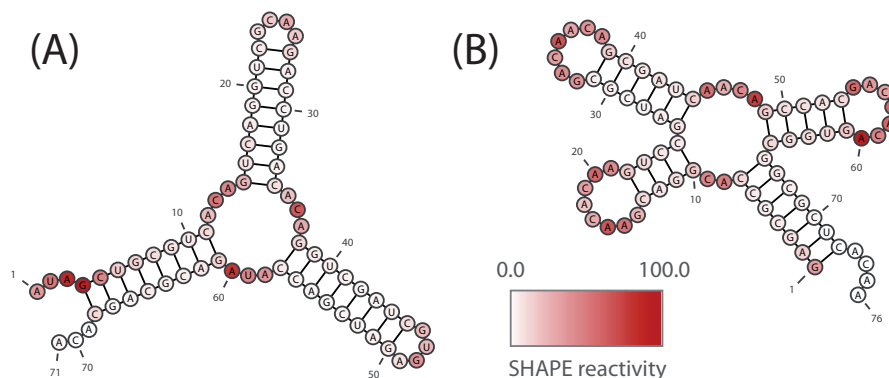
Figure 1. Relative SHAPE reactivities of the two experimentally tested RNA molecules overlaid over their respective target secondary structures, the three-arm junction (A) and the tRNA-like four-arm junction (B).

secondary structure cannot be determined directly with this method alone. Nonetheless, the measured reactivities are consistent with the target secondary structures.

4 Discussion

Despite the simplifications used in the free energy model and the negative design term, the method seems to work quite well and we were able to experimentally test our method for two example targets. We hope to extend the method to more detailed structural models of nucleic acids in the future.

References

1. I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, P. Schuster, *Fast Folding and Comparison of RNA Secondary Structures*, Monatsh. Chem. **125**, 167–188, 1994.

2. D. H. Mathews, D. H. Turner, *Prediction of RNA secondary structure by free energy minimization*, Curr. Opin. Struct. Biol. **16**, 270–278, 2006.

3. M. Andronescu, A. P. Fejes, F. Hutter, H. H. Hoos, A. Condon, *A New Algorithm for RNA Secondary Structure Design*, J. Mol. Biol. **336**, 607–624, 2004.

4. A. Busch, R. Backofen, *INFO-RNA—a fast approach to inverse RNA folding*, Bioinformatics **22**, 1823–1831, 2006.

5. J. N. Zadeh, B. R. Wolfe, N. A. Pierce, *Nucleic acid sequence design via efficient ensemble defect optimization*, J. Comput. Chem. **32**, 439–452, 2011.

6. J. L. Knight, C. L. Brooks III, *λ-Dynamics free energy simulation methods*, J. Comput. Chem. **30**, 1692–1700, 2009.

7. I. Lafontaine, R. Lavery, *Optimization of Nucleic Acid Sequences*, Biophys. J. **79**, 680–685, 2000.

8. K. A. Wilkinson, E. J. Merino, K. M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution*, Nat. Protoc. **1**, 1610–1616, 2006.

Cellular Events in Anticancer and Anti AIDS Therapy Investigated by Computational Methods

**Vo Cam Quy, Trung Hai Nguyen, Emiliano Ippoliti,
Giulia Rossetti, and Paolo Carloni**

German Research School for Simulation Sciences, 52425 Jülich, Germany
E-mail: p.carloni@grs-sim.de

Cancer and HIV/AIDS are among the worst fatal diseases world wide. Fighting against these diseases requires overcoming many challenging problems, among which drug resistance represents a major cause of treatment failure. Understanding cellular mechanisms responsible for drug resistance is essential for the development of new drugs with improved efficacy. In this article, we introduce the computational studies on the cellular events leading to drug resistance of cancer and HIV/AIDS diseases. In particular, our studies focus on how the widely used anticancer drug cisplatin interacts with its cellular transport proteins, which have been found to play a key role in resistance of the drug. In case of HIV-1 we investigate the structural characteristic of viral-cellular protein complexes, which represent a novel target for new drug development strategies to avoid drug resistance caused by rapid mutation of viral proteins.

1   Introduction

Cisplatin is the one of most widely used anticancer drugs[1]. It suffers, however, from several drawbacks such as side effects and drug resistance, which severely limit its efficacy[3]. Cellular mechanisms behind resistance of cisplatin are multifactorial and in general not very well understood[4]. It is known, however, that reduced cellular accumulation of the drug is a common and distinctive feature of resistant cells[4]. It is, therefore, important to understand the cellular influx and efflux of cisplatin. The copper transporter *Ctr1* has been found to play an important role in cellular uptake of cisplatin, and its methionine (Met)-rich N-term domain has been identified as a putative cisplatin binding site[4]. The reaction between cisplatin with one of the conserved Met-rich motifs, called *Mets7*, was characterized experimentally by the electrospray mass (ESI-MS), NMR, circular dichroism (CD)[5] and EXAFS (Extended X-ray Absorption Fine Structure)[2] spectroscopy methods. The study showed that cisplatin progressively loses all ammine and chloride ligands and ultimately coordinates to *Mets7* in a naked form[5]. While experimental methods can provide only information of the Pt coordination as well as the secondary-structure types of the *Mets7* peptide, an atomic description of the cisplatin-*Mets7* adducts can be provided by means of molecular simulations.

HIV is a retrovirus which mutates rapidly its genome and becomes capable of eluding the action of drugs targeting the current viral protein targets. A very promising strategy to overcome viral hyper-variability is to design compounds capable of interfering with the interactions between viral proteins and human factors, since the latter are not expected to mutate normally. Transcriptional activation is an essential step in the replication cycle of HIV. Since Tat (Trans-Activator of Transcription) plays a key role in viral transcription, and exerts its function by binding viral DNA as well as several cellular proteins, it is considered a target of interest to develop innovative anti-HIV therapies. The process of trans-activation

is complex and requires the participation of several human co-factors (CyclinT1, CDK9, p/CAF and others). The lack of detailed atomistic structural information of Tat-host protein complexes requires a molecular modeling strategy, validated through molecular biology experiments. In this regard, we aim to build a structural model of Tat-p/CAF. This will help us in the understanding of possible new drug targeting HIV-1 Tat-induced transcription.

2 Methods

To study the structural and spectroscopic properties of cisplatin-*Mets7* adduct we use hybrid quantum mechanical / molecular mechanics (QM/MM) simulations[6] using CPMD program,[15] the force matching technique[7] and classical molecular dynamics (MD) simulations using NAMD package[12]. Two models have been simulated, in which the fourth ligand bound to the Pt is either chloride or a water molecule. The calculated spectroscopic properties (NMR[9], EXAFS[8] and CD[10]) will be compared with experimental data to validate our model.

To study Tat_{full_length} formation with one of the cellular partners, p/CAF, we performed several multi-conformers dockings using HADDOCK 2.1[11] together with MD simulations. The MD calculations were performed with NAMD package by using AMBER force field ff99SB for Tat_{full_length} AcK50 and Tat_{full_length}.p/CAF docking model. The acetylated lysine topology parameter was adapted according to Pantano *et al.*[13].

3 Results

3.1 Cisplatin-Mets7

25ps long QM/MM simulatons have been performed for the cisplatin-Mets7 adduct models. The EXAFS spectra at the Pt edge have been calculated and showed good agreement with the experimental spectrum. This indicates that the structural and dynamical properties of the platinum coordination are basically reproduced by our QM/MM simulations. The calculated ^{1}H and ^{13}C NMR chemical shifts generally agree with the experimental values within error bars. To study the global conformation of the whole platinated peptides, 0.8 μs classical MD simulations has been carried out based on the force field parameters obtained from the force matching procedure[7]. During the classical MD simulations, the models adopt very rigid conformations. The calculated CD spectra show qualitative agreement with experiment.

3.2 Tat-P/CAF

Tat does not have prominent secondary structure elements in the absence of interacting partners, which indicates its flexible nature in the unbound conditions[14]. Therefore, after minimization, all the principal conformations visited during the molecular dynamics were identified by a clustering analysis. Representatives of each cluster have been used as the input for Tat_{full_length} in the docking protocol. By multiple conformation docking and MD simulations, we have predicted the structural determinants of Tat in complex with one of the host cell cognate proteins, the transcriptional co-activator and histone-acetyltransferase

p/CAF. All of the computational results are in good agreement with experimental data[11, 13]. The understanding at atomic level may help the design of ligands interfering with the Tat function.

Acknowledgments

References

1. R. A. Alderden, A. D. Hall and T. W. Hambley, *The discovery and development of cisplatin*, J. Chem. Educ. **83**, 728–734, 2006.
2. R. A. Arnesano *et al.*, Unpublished data.
3. P. A. Andrews and S. B. Howell, *Cellular pharmacology of cisplatin: perspectives on mechanisms of acquired resistance*, Cancer Cells **2**, 35–43, 1990.
4. M. D. Hall, M. Okabe, D. W. Shen, X. J. Liang and M. M. Gottesman, *The Role of Cellular Accumulation in Determining Sensitivity to Platinum-Based Chemotherapy*, Annu. Rev. Pharmacol. Toxicol. **48**, 495–535, 2008.
5. F. Arnesano, S. Scintilla and G. Natile, *Interaction between platinum complexes and a methionine motif found in copper transport proteins*, Angew. Chem. Int. Ed. Engl. **46**, 9062–9064, 2007.
6. A. Laio, J. VandeVondele and U. Röthlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys. **116**, 6941–6948, 2002 and references therein.
7. P. Maurer, A. Laio, H. W. Hugosson, M. C. Colombo and U. Röthlisberger, *Automated parametrization of biomolecular force fields from quantum mechanics/molecular mechanics (QM/MM) simulations through force matching*, J. Chem. Theory Comput. **3**, 628–639, 2007.
8. J. J. Rehr, R. C. Albers and S. I. Zabinsky, *High-order multiple-scattering calculations of x-ray-absorption fine structure*, Phys. Rev. Lett. **69**, 3397–3400, 1992.
9. D. Sebastiani and M. Parrinello, *A new ab-initio spproach for NMR shemical shifts in periodic systems*, J. Phys. Chem. A **105**, 1951–1958, 2001.
10. B. M. Bulheller and J. D. Hirst, *DichroCalc–circular and linear dichroism online*, Bioinformatics **25**, 539–540, 2009.
11. M. R. Machado, P. D. Dans and S. Pantano, *Isoform-specific determinants in the HP1 binding to histone 3: insights from molecular simulations*, Amino Acids **38**, 1571–1581, 2010.
12. J. C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *Scalable molecular dynamics with NAMD*, J. Comp. Chem. **26**, 1781, 2005.
13. T. H. Tahirov, N. D. Babayeva, K. Varzavand, J. J. Cooper, S. C. Sedore and D. H. Price, *Crystal structure of HIV-1 Tat complexed with human P-TEFb*, Nature **465**, 747–751, 2010.

14. C. Dominguez, R. Boelens and A. M. J. J. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information*, J. Am. Chem. Soc. **125**, 1731–1737, 2003.

15. IBM Corp 1990-2008, MPI für Festkörperforschung Stuttgart 1997-2001, *CPMD: 3.13.2 edition*, `http://www.cpmd.org/` (2009).

Generalized Isobaric-isothermal Ensemble Algorithms

Yoshiharu Mori[1] and Yuko Okamoto[1,2,3]

[1] Department of Physics, Graduate School of Science, Nagoya University
Nagoya, Aichi 464-8602, Japan
E-mail: ymori@tb.phys.nagoya-u.ac.jp

[2] Structural Biology Research Center, Graduate School of Science, Nagoya University
Nagoya, Aichi 464-8602, Japan

[3] Center for Computational Science, Graduate School of Engineering, Nagoya University
Nagoya, Aichi 464-8603, Japan
E-mail: okamoto@phys.nagoya-u.ac.jp

Generalized-ensemble algorithms for isobaric-isothermal molecular simulations are described. Extensions of multicanonical algorithm, replica-exchange method, and simulated tempering in the canonical ensemble to the isobaric-isothermal ensemble are explained.

1 Introduction

Monte Carlo (MC) and molecular dynamics (MD) simulations of complex systems such as biomolecular systems are greatly hampered by the multiple-minima problem[1,2]. To overcome this difficulty, we have been advocating the uses of the generalized-ensemble algorithms (for reviews, see, e.g., Ref. 3, 4). Three well-known generalized-ensemble algorithms are the multicanonical algorithm (MUCA)[5] (the MD version was developed in Ref. 6, 7), replica-exchange method (REM)[8] (the MD version, which is referred to as REMD, was developed in Ref. 9), and simulated tempering (ST)[10,11]. These methods can be used for obtaining accurate physical quantities in the canonical ensemble. Among them, REM (particularly REMD) is often used because the weight factor is *a priori* known (i.e., the Boltzmann factor), while those for MUCA and ST have to be determined by preliminary simulations before the production runs.

Multidimensional (or multivariable) extensions of the original generalized-ensemble algorithms have been developed in many ways (see the Ref. in 4) and recently the general formulations have been given in Ref. 12.

In this article, we describe three generalized-ensemble algorithms for the NPT ensemble. Namely, extensions of MUCA, REM, and ST from the canonical ensemble to the isobaric-isothermal ensemble will be given. They are the multibaric-multithermal (MU-BATH) algorithm[13], REM[14–17], and ST[18] for the NPT ensemble.

2 Methods

We first briefly review the generalized-ensemble algorithms for isobaric-isothermal molecular simulations. Let us consider a physical system that consists of N atoms and that is in a box of a finite volume V. The states of the system are specified by coordinates $r \equiv \{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}$ and momenta $p \equiv \{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_N\}$ of the atoms and volume V of the box. The potential energy $E(r, V)$ for the system is a function of r and V.

We first describe MC simulation algorithms for MUCA, REM, and ST in the NPT ensemble. In these cases, momenta of atoms do not have to be considered. To make a system an equilibrium state, the detailed balance condition is imposed and a transition probability $w(X \to X')$ from an old state X to a new state X' can be given by the Metropolis criterion[19].

In MUBATH simulations, we introduce a function $\mathcal{H}(E, V)$ and use a weight factor $W_{\mathrm{mbt}}(E, V) \equiv \exp\left[-\beta_0 \mathcal{H}(E, V)\right]$ so that the distribution function $f_{\mathrm{mbt}}(E, V)$ of E and V may be uniform:

$$f_{\mathrm{mbt}}(E, V) \propto n(E, V) W_{\mathrm{mbt}}(E, V) = \text{constant}, \tag{1}$$

where β_0 is an arbitrary inverse reference temperature defined as $\beta_0 = 1/k_{\mathrm{B}} T_0$ (k_{B} is the Boltzmann constant) and $n(E, V)$ is the density of states.

To perform MUBATH MC simulations, the trial moves are generated in the same way as in the usual constant NPT MC simulations[20] and the transition probability from $X \equiv \{s, V\}$ to $X' \equiv \{s', V'\}$ is given by[13]

$$w_{\mathrm{mbt}}(X \to X') = \min\left[1, \exp(-\Delta_{\mathrm{mbt}})\right], \tag{2}$$

where

$$\Delta_{\mathrm{mbt}} = \beta_0 \left\{ \mathcal{H}\left[E(s', V'), V'\right] - \mathcal{H}\left[E(s, V), V\right] - N k_{\mathrm{B}} T_0 \ln(V'/V) \right\}, \tag{3}$$

and $s = \{s_1, s_2, \cdots, s_N\}$ is the scaled coordinates defined by $s_i = V^{-1/3} r_i$ ($i = 1, 2, \cdots, N$). Here, we are assuming the box is a cube of side $V^{-1/3}$.

In REM simulations, we prepare a system that consists of $M_T \times M_P$ non-interacting replicas of the original system, where M_T and M_P are the number of temperature and pressure values used in the simulation, respectively. The replicas are specified by labels i ($i = 1, 2, \cdots, M_T \times M_P$), temperature by m_t ($m_t = 1, 2, \cdots, M_T$), and pressure by m_p ($m_p = 1, 2, \cdots, M_P$).

To perform REM MC (REMC) simulations, we carry out the following two steps alternately: (1) perform a usual constant NPT simulation in each replica at assigned temperature and pressure and (2) try to exchange the replicas. If the temperature (specified by m_t and n_t) and pressure (specified by m_p and n_p) between the replicas are exchanged, the transition probability from $X \equiv \{\cdots, (s^{[i]}, V^{[i]}; T_{m_t}, P_{m_p}), \cdots, (s^{[j]}, V^{[j]}; T_{n_t}, P_{n_p}), \cdots\}$ to $X' \equiv \{\cdots, (s^{[i]}, V^{[i]}; T_{n_t}, P_{n_p}), \cdots, (s^{[j]}, V^{[j]}; T_{m_t}, P_{m_p}), \cdots\}$ at the trial is given by[15,16]

$$w_{\mathrm{rem}}(X \to X') = \min\left[1, \exp(-\Delta_{\mathrm{rem}})\right], \tag{4}$$

where

$$\begin{aligned} \Delta_{\mathrm{rem}} &= (\beta_{m_t} - \beta_{n_t}) \left[E(s^{[j]}, V^{[j]}) - E(s^{[i]}, V^{[i]}) \right] \\ &\quad + (\beta_{m_t} P_{m_p} - \beta_{n_t} P_{n_p}) \left(V^{[j]} - V^{[i]} \right). \end{aligned} \tag{5}$$

In ST simulations, we introduce a function $g(T, P)$ and use a weight factor $W_{\mathrm{st}}(E, V; T, P) \equiv \exp[-\beta(E + PV) + g(T, P)]$ so that the distribution function $f_{\mathrm{st}}(T, P)$ of T and P may be uniform:

$$f_{\mathrm{st}}(T, P) \propto \int_0^\infty dV \int_V dr \, W_{\mathrm{st}}[E(r, V), V; T, P] = \text{constant}. \tag{6}$$

From Eq. (6), it is found that $g(T, P)$ is formally given by

$$g(T, P) = -\ln \left\{ \int_0^\infty dV \int_V dr \exp\left[-\beta\left(E(r, V) + PV\right)\right] \right\}, \qquad (7)$$

and the function is the dimensionless Gibbs free energy except for a constant.

To perform ST MC simulations, we carry out the following two steps alternately: (1) perform a usual constant NPT simulation and (2) try to update the temperature and pressure. The transition probability from $X \equiv \{s, V; T, P\}$ to $X' \equiv \{s, V; T', P'\}$ for this trial is given by[18]

$$w_{\text{st}}(X \to X') = \min\left[1, \exp(-\Delta_{\text{st}})\right], \qquad (8)$$

where

$$\Delta_{\text{st}} = (\beta' - \beta)E(s, V) + (\beta'P' - \beta P)V - \left[g(T', P') - g(T, P)\right]. \qquad (9)$$

For MD simulations with MUCA, REM, and ST in the NPT ensemble, the actual formulations depend on constant temperature and pressure algorithms. Here, we employ the MD methods with the Martyna-Tobias-Klein (MTK) algorithm[21], whose equations of motion follow Nosé[22] and Hoover[23] for the thermostat and Andersen[24] for the barostat.

While the weight factors in REM simulations are *a priori* known, those in MUBATH and ST simulations have to be determined before the simulations.

3   Results and Discussion

In order to verify that the generalized-ensemble algorithms discussed above can be effective for conformational sampling and give the same results, we performed MD simulations with the three generalized-ensemble algorithms. We used a system of an alanine dipeptide in 73 surrounding water molecules and the system was placed in a cubic cell with periodic boundary conditions. Both of the backbone dihedral angles ϕ and ψ of the peptide were initially set to $180°$.

First, we performed the two-dimensional REMD simulation. The simulation time was set to 2.0 ns. We used the following six temperature $(T_1, \cdots, T_6)$ and four pressure $(P_1, \cdots, P_4)$ values: 280, 305, 332, 362, 395, and 430 K for temperature and 0.1, 65, 150, and 250 MPa for pressure. At the replica-exchange trial, either exchanging temperature (T-exchange) or exchanging pressure (P-exchange) was chosen randomly.

Fig. 1(a) shows the time series of T and P in one of the replicas and Fig. 1(b) shows the time series of the label of the replicas in the simulation at 430 K and 0.1 MPa. From these figures, it is found that random walks in T-P space were realized in each of the replicas.

We then performed 24 MUBATH simulations of 2.0 ns, where the total simulation time was 48 ns so that it is equal to that in the REMD simulation. In each of the simulations, different initial velocities were given.

Figures 1(c) and (d) show the probability distributions of E and V from the MUBATH simulations. Fig. 1(a) shows the time series of T and P in one of the replicas and Fig. 1(b) shows the time series of the label of the replicas in the simulation at 430 K and 0.1 MPa. From Fig. 1(c), it is found that the MUBATH simulations gave a uniform distribution in the range where the density of states was obtained accurately in the REMD simulation and
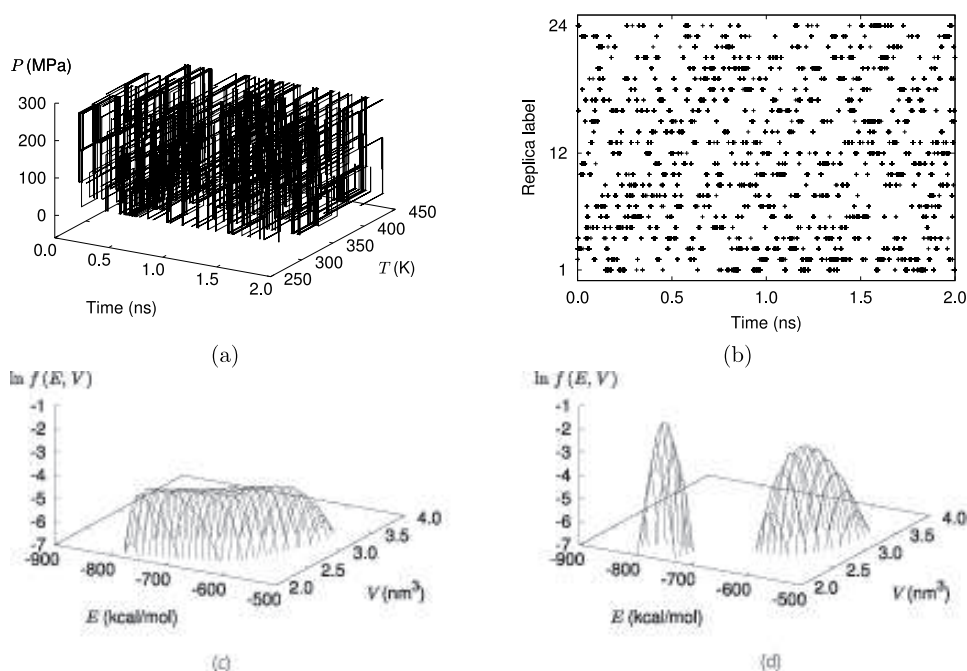
Figure 1. Results of the two-dimensional REMD simulation: (a) the time series of T and P in one of the replicas and (b) the time series of the replica label at 430 K and 0.1 MPa and logarithm of the probability distributions f of E and V (c) in the MUBATH simulations and (d) for the NPT ensemble at 280 K and 250 MPa (left) and at 430 K and 0.1 MPa (right).
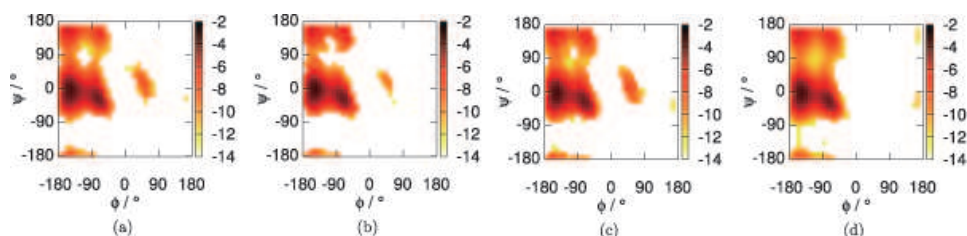

Figure 2. Contour maps of probability distribution of the backbone dihedral angles ϕ and ψ in the simulations with (a) REMD, (b) MUBATH, and (c) ST and (d) in the conventional isobaric-isothermal simulations. In these figures, the probability distributions at 298 K and 0.1 MPa are plotted in logarithmic scale.

that the distribution of the MUBATH simulations was much wider than the ones for the NPT ensemble (see Fig. 1(d)).

Twenty-four ST simulations of 2.0 ns were also carried out, which gave us the same number of sampled data as in the REMD and MUBATH simulations. In each simulation, different initial velocities were given. We used the same temperature and pressure values as in the REMD simulation.

Fig. 2 shows the probability distributions of the backbone dihedral angles at 298 K and

0.1 MPa in all the simulations. Compared with the simulations with REMD, MUBATH, and ST, we also performed 24 conventional isobaric-isothermal simulations of 2.0 ns at 298 K and 0.1 MPa with different initial velocities. The dihedral angle distributions in the simulations with the generalized-ensemble algorithms had a small peak in $0° \leq \phi \leq 90°$ and $-90° \leq \psi \leq 90°$ although there was no peak in the range in the distribution of the conventional simulations. All the simulations with the three generalized-ensemble algorithms were able to reproduce the distribution obtained previously[25].

4 Conclusions

In this article, we presented the extensions of MUCA, REM, and ST for isobaric-isothermal molecular simulations. These algorithms can be effective methods for conformational sampling and give accurate physical quantities in the isobaric-isothermal ensemble. Therefore, one can use the generalized-ensemble algorithms to study temperature and pressure effects on complex systems such as biomolecular systems.

Acknowledgments

References

1. H.A. Scheraga, J. Lee, J. Pillardy, Y.-J. Ye, A. Liwo, and D. Ripoll, J. Glob. Optimiz. **15**, 235–260, 1999.
2. D.J. Wales and H.A. Scheraga, Science **285**, 1366–1372, 1999.
3. U.H.E. Hansmann and Y. Okamoto, Curr. Opin. Struct. Biol. **9**, 177–183, 1999.
4. A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96–123, 2001.
5. B.A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249–253, 1991; Phys. Rev. Lett. **68**, 9–12, 1992.
6. U.H.E. Hansmann, Y. Okamoto, and F. Eisenmenger, Chem. Phys. Lett. **259**, 321–330, 1996.
7. N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817–824, 1997.
8. K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604–1608, 1996.
9. Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141–151, 1999.
10. A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, and P.N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776-1783, 1992.
11. E. Marinari and G. Parisi, Europhys. Lett. **19**, 451–458, 1992.
12. A. Mitsutake and Y. Okamoto, Phys. Rev. E **79**, 047701 (4 pages), 2009; J. Chem. Phys. **130**, 214105 (14 pages), 2009.
13. H. Okumura and Y. Okamoto, Chem. Phys. Lett. **383**, 391-396, 2004; Chem. Phys. Lett. **391**, 248–253, 2004.

14. T. Nishikawa, H. Ohtsuka, Y. Sugita, M. Mikami, and Y. Okamoto, Prog. Theor. Phys. Suppl. **138**, 270–271, 2000.

15. T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami, Chem. Phys. Lett. **335**, 435–439, 2001.

16. Y. Sugita and Y. Okamoto, in *Lecture Notes in Computational Science and Engineering*, T. Schlick and H.H. Gan, eds., Springer-Verlag, Berlin, pp. 303-331, 2002; available at arXiv:cond-mat/0102296 [cond-mat.stat-mech].

17. D. Paschek and A. E. García, Phys. Rev. Lett. **93**, 238105 (4 pages), 2004.

18. Y. Mori and Y. Okamoto, J. Phys. Soc. Jpn. **79**, 074003 (5 pages), 2010.

19. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087–1092, 1953.

20. I. R. McDonald, Mol. Phys. **23**, 41–58, 1972.

21. G.J. Martyna, D.J. Tobias, and M.L. Klein, J. Chem. Phys. **101**, 4177–4189, 1994.

22. S. Nosé, Mol. Phys. **52**, 255–268, 1984; J. Chem. Phys. **81**, 511–519, 1984.

23. W.G. Hoover, Phys. Rev. A **31**, 1695–1697, 1985.

24. H.C. Andersen, J. Chem. Phys. **72**, 2384–2393, 1980.

25. H. Okumura and Y. Okamoto, J. Phys. Chem. B **112**, 12038–12049, 2008.

Secondary Structure Propensities of the Amyloid β-Peptide Aβ_{1-42} as Influenced by pH and a D-Peptide

Olujide O. Olubiyi[1,2] and Birgit Strodel[1]

[1] Institute of Complex Systems: Structural Biochemistry, Forschungszentrum Jülich
52425 Jülich, Germany
E-mail: {o.olubiyi, b.strodel}@fz-juelich.de

[2] German Research School for Simulation Sciences GmbH, 52425 Jülich, Germany

We investigate influences on the secondary structure of one of the major alloforms of the Alzheimers peptide, Aβ_{1-42}. Specifically, we study how the pH and binding of D3, a 12-residue D-peptide with demonstrated anti-amyloid effects, affect Aβ_{1-42} folding. We demonstrate that, under slightly acidic conditions, protonation of the three histidine residues in Aβ_{1-42} promotes the formation of β-sheets via a reduction in electrostatic repulsion between the two terminal domains. Our studies further reveal crucial aspects of Aβ_{1-42} important for its interaction with D3, involving four negatively charged residues in the N-terminal section of Aβ_{1-42}. The binding of D3 was found to induce large conformational changes in the amyloid peptide with a reduction in β-sheet being the most significant effect possibly explaining the observed aggregation-inhibiting properties of the D-peptide.

1 Introduction

Protein folding abnormalities are an important class of pathogenic causes of diseases like Alzheimer's and Parkinson's disease. In Alzheimer's disease (AD) post-translational cleavage products of the amyloid precursor protein (APP) have been identified as a clinical hallmark in disease development and progression. Of the APP cleavage products Aβ_{1-42} is recognized as the most important alloform based primarily on roles in eliciting neurotoxicity. Conformational transition to a predominantly β-sheet of the extracellular Aβ peptide resulting in its aggregation into water-soluble oligomers is believed to be crucial in the initiation of Alzheimer's dementia. The peptide has been shown to sample a wide range of conformations from β-sheets to random coil and helical structures. The aggregation of Aβ monomers into toxic β-sheet-rich oligomer structures is believed to depend to a significant extent on the sampled conformational state of the monomer and factors influencing it, including the pH, metal ions, lipid membranes, and the presence of preformed oligomers[1,2].

The molecular events surrounding its causative role in AD make Aβ a viable target for drug discovery purpose. Different therapeutic strategies have been employed exploiting conformational aspects of Aβ including peptidic inhibitors, which can prevent the formation of neurotoxic aggregates. Peptide drugs, however, suffer from a significant drawback that is the fast rate at which they are cleared from circulation by endogenous peptidases. Approaches employed in circumventing this disadvantage include D-peptides. Recently a D-enantiomeric peptide, D3 (a 12-residue arginine-rich peptide), with anti-amyloid aggregation properties, was discovered by mirror-image phage display[3]. In AD mice models, D3 was not only demonstrated to improve the pathology and behavior, it was also shown to be orally bioavailable[4].

In this study we use multiple microsecond-long MD simulations in explicit water to study various influences on the secondary structure of the Aβ peptide. We present a perspective on the intramolecular effects of the histidine protonation state on the sampled structures of Aβ_{1-42} and how this may be the basis for its increased aggregation kinetics at acidic pH. Using the insights afforded by Aβ conformational studies we proceed to investigate the reported anti-amyloid properties of D3. To unravel D3's mechanism of amyloid aggregation inhibition we employ a combination of a global optimization method and MD simulations.

2   Methods

Structural Models Initial coordinates for the D3 peptide were generated using the Dundee PRODRG2 server. The solution NMR structures PDB 1Z0Q for Aβ_{1-42} obtained in apolar solvent was employed as starting configuration for the MD simulations. If not stated otherwise, histidine residues were modeled as neutral with only the delta nitrogen protonated. The N-termini and C-termini were respectively chosen to be protonated and deprotonated in both D3 and Aβ peptides to mimic the physiological states at pH 7.4.

Molecular dynamics simulations Multiple MD simulations were performed beginning with a 100 ns simulation of D3 using the GROMOS ffG43a2 force field[5]. The peptide was centered in a cubic simulation box with a 1 nm distance allowed between the peptide and the edges of the box treated with periodic boundary conditions. The Particle-Mesh Ewald method was employed for treating long-range electrostatics with a 1.4 nm cut-off used for calculating short-range forces. The box was solvated using the SPC explicit water model, and Na$^+$ and Cl$^-$ ions added to obtain a NaCl concentration of 150 mM and achieve charge neutrality. The solvated peptide was then minimized and subsequently equilibrated in a 1 ns MD simulation with position restraints on all non-hydrogen atoms of D3. The restraints were then turned off and the 100 ns production run performed at 300 K in an NPT ensemble. The procedure followed for the preparation of the MD simulation of D3 was also employed for the MD simulations of Aβ_{1-42}, followed by 1 μs production runs. To investigate the effect of histidine protonation on Aβ_{1-42} folding we performed another 1 μs MD simulation with the three histidine residues of Aβ_{1-42} (His6, His13, His14) protonated on both imidazolyl nitrogen atoms. In addition, from the 1 μs MD simulation of Aβ_{1-42} with neutral histidine residues three frames dumped at 713 ns, 817 ns and 888 ns were each subjected to further 335 ns MD simulations (totaling $\approx$ 1 μs). These MD simulations were used as a reference for quantifying the effect of D3 binding on the secondary structural features of Aβ_{1-42}. All MD simulations were performed with GROMACS[6].

Basin-Hopping global optimization The basin-hopping (BH) approach to global optimization[7] is analogous in idea to the Monte Carlo-minimization approach[8]. Moves are proposed by perturbing the current geometry, and are accepted or rejected based upon the energy difference between the local minimum obtained by minimization from the instantaneous configuration and the previous minimum in the chain. Large steps can be taken to sample the energy landscape, since the objective is to step between local minima. Furthermore, there is no need to maintain detailed balance when taking steps, because
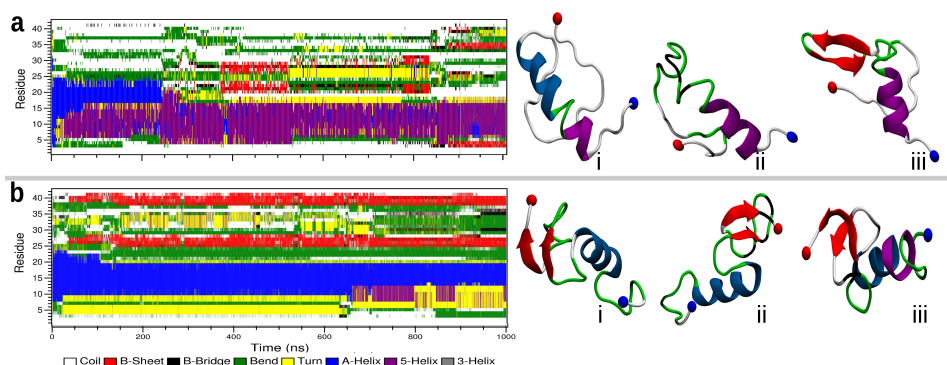
Figure 1. Left: DSSP plots for secondary structure transitions in $A\beta_{1-42}$ with (a) neutral and (b) protonated histitdine residues during 1 μs MD simulations. Right: The $A\beta_{1-42}$ structures for the centers of the three most populated clusters sampled in the MD simulations. The peptide is colored according to secondary structure: red for β-sheet, blue for α-helix, purple for π-helix, green for bend, yellow for turn and white for coil structure. The N- and C-termini are respectively shown in blue and red beads.

the BH approach attempts to locate the global potential energy minimum and is not intended to sample thermodynamic properties. The BH algorithm has been implemented in the GMIN program[9]. We used the oligomer-generation procedure in GMIN[10] to generate 6000 $A\beta_{1-42}$/D3 complexes from three different $A\beta_{1-42}$ structures and the most favored structure of D3. After their initial generation, the complexes were optimized using 500 BH steps with dihedral angle moves and small rigid body rotations and translations applied to D3. All the systems were treated with the parameters from the CHARMM22 force field. From the 6000 complexes 100 low-energy structures were selected for subsequent MD simulations with care taken to ensure that the three $A\beta_{1-42}$ starting structures employed for the BH runs were equally represented. The MD production runs were performed for 10 ns (totaling 1 μs) in explicit solvent using the GROMOS ffG43a2 force field[5] following a 1 ns equilibration run.

3 Results and Discussion

The effect of protonating both imidazolyl nitrogen atoms is a change of the overall charge of $A\beta_{1-42}$ from -3 to zero. We therefore use $A\beta_{1-42}^{3-}$ and $A\beta_{1-42}^{0}$ in the following to refer to $A\beta_{1-42}$ with neutral and positively charged (also denoted as protonated) histidine residues, respectively. The secondary structure transitions in each of the two 1 μs MD simulations were monitored using the DSSP program[11]. The resulting plots are shown in Fig. 1 along with representative structures for the three most populated clusters as obtained from conformational clustering[12]. We observe a stabilization of the N-terminal α-helix and a higher amount of β-sheets for $A\beta_{1-42}^{0}$ compared to $A\beta_{1-42}^{3-}$. Analysis of the underlying molecular basis for these observations reveals that the protonation of the histdine residues facilitates contacts within the otherwise negatively charged N-terminal segment via a reduction in the local net charge of the first 23 residues, which also allows contacts with the hydrophobic C-terminal. In $A\beta_{1-42}^{3-}$ the N- and C-terminal segments avoid interacting
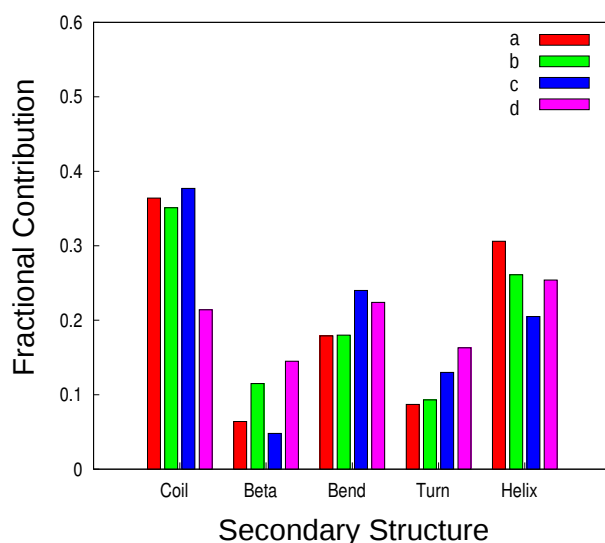
Figure 2. Fractional secondary structure contributions obtained for (a) $A\beta^{3-}_{1-42}$ from a 1 μs MD simulation, (b) three 335 ns MD simulations using coordinates of $A\beta^{3-}_{1-42}$ dumped at 713 ns, 817 ns and 888 ns, (c) D3-bound $A\beta^{3-}_{1-42}$ from 100 individual 10 ns MD simulations, and (d) $A\beta^{0}_{1-42}$ from a 1 μs MD simulation.

with each other, which is likely a result of charge incompatibility between the hydrophobic C-terminal and the hydrophilic N-terminal. The hydrophobic interactions between the two terminal sections in $A\beta^{0}_{1-42}$ induce a significant increase in β-sheet formation. This is an important driving force for the fast aggregation kinetics observed for the amyloid peptide at acidic pH[2], which should be further enhanced from reduction in electrostatic repulsion between $A\beta$ molecules. This result agrees with the observation that electrostatic repulsion prevents $A\beta$ aggregation[1]. The effect of a pronounced increase in β-sheet content and a reduction of coil structures due to the protonation of the histidine residues is also very well displayed by the fractional secondary structure units obtained from the 1 μs MD simulations of $A\beta^{3-}_{1-42}$ and $A\beta^{0}_{1-42}$ in Fig. 2.

From the 100 ns MD simulation for the D3 peptide we observed that the conformational space is dominated by an extended conformation, which results from an attempt to minimize electrostatic repulsion between its five arginine residues. We employed this structure for our subsequent simulations investigating D3 binding to $A\beta_{1-42}$ (with neutral histidine residues). For $A\beta_{1-42}$ we selected three different structures from the 1 μs MD simulation making sure that helical, β-strand and coil secondary structures were adequately represented. For the 10 ns MD simulations, which were performed for 100 $A\beta_{1-42}$/D3 complexes, the Coulombic and Lennard-Jones interaction energies were decomposed into residual contributions based on the interaction of each $A\beta_{1-42}$ amino acid residue with D3. The result of this analysis, averaged over 10 ns and 100 simulations, revealed that the interaction is mediated by strong electrostatic attraction between the arginine-rich D3 and $A\beta_{1-42}$ since the strongest interactions are obtained for the four negatively charged aspartic and glutamic residues in the N-terminal segment of $A\beta_{1-42}$. We then investigated

the effect of D3 binding on the secondary structure of the $A\beta_{1-42}$ monomer (Fig. 2). We calculated the fractional secondary structure units for $A\beta$ in the $A\beta_{1-42}$/D3 complexes for each of the 10 ns MD simulations and averaged these values over the 100 systems. We compared the results with the secondary structure contributions sampled in three 335 ns MD simulations, which were initiated with the $A\beta_{1-42}$ structures used for the generation of $A\beta_{1-42}$/D3 complexes. The overall effect of D3 on $A\beta_{1-42}$ involves a significant decrease in β-sheet (58.3%) and helical (21.5%) contents and an increase (19.7%) in total coil structures (coil, turn and bend) suggesting that the binding of D3 to $A\beta_{1-42}$ induces conversion of β-sheets and helices into random coil states. The destruction of β-sheet in $A\beta_{1-42}$ monomer probably provides an explanation for the amyloid aggregation inhibition observed both in vitro and in vivo [49].

Acknowledgments

References

1. M. Guo, P.M. Gorman, M. Rico, A. Chakrabartty, and D.V. Laurents, *Charge substitution shows that repulsive electrostatic interactions impede the oligomerization of Alzheimer amyloid peptides.*, FEBS Lett., **579**, 3574–3578, 2005.

2. Y. Su and P.T. Chang, *Acidic pH promotes the formation of toxic fibrils from beta-amyloid peptide.*, Brain Res., **893**, 287–291, 2001.

3. T. van Groen, I. Kadish, K. Wiesehan, S. A. Funke, and D. Willbold, *In vitro and in vivo staining characteristics of small, fluorescent, A42 binding D-enantiomeric peptides in transgenic AD mouse models*, ChemMedChem, **4**, 276–282, 2009.

4. S. A. Funke, T. van Groen, I. Kadish, D. Bartnik, L. Nagel-Steger, O. Brener, T. Sehl, R. Batra-Safferling, C. Moriscot, G. Schoehn, A.H.C. Horn, A. Müller-Schiffmann, C. Korth, H. Sticht, and D. Willbold, *Oral Treatment with the D-Enantiomeric Peptide D3 Improves Pathology and Behavior of Alzheimers disease Transgenic Mice*, ACS Chem. Neurosci., **1**, 639–648, 2010.

5. L. Schuler, X. Daura, and W. van Gunsteren, *An improved Gromos96 force field for aliphatic hydrocarbons in the condensed phase.*, J. Comput. Chem., **22**, 1205–1218, 2001.

6. B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theor. Comput., **4**, 435–447, 2008.

7. D. J. Wales and J. P. K. Doye, *Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms*, J. Phys. Chem. A, **101**, 5111–5116, 1997.

8. Z. Li and H. A. Scheraga, *Monte Carlo-minimization approach to the multiple-minima problem in protein folding*, Proc. Natl. Acad. Sci. USA, **84**, 6611–6615, 1987.

9. D. J. Wales, *GMIN: A program for basin-hopping global optimisation*, http://www-wales.ch.cam.ac.uk/software.html.

10. B. Strodel, J.W.L. Lee, C.S. Whittleston, and D.J. Wales, *Transmembrane structures for Alzheimer's Aβ_{1-42} oligomers.*, J. Am. Chem. Soc., **132**, 13300–13312, 2010.

11. W. Kabsch and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*, Biopolymers, **22**, 2577–2637, 1983.

12. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. van Gunsteren, and A. Mark, *Peptide folding: when simulation meets experiment.*, Angew. Chem. Int. Ed., **38**, 236–240, 1999.

Coarse Grained Simulation of Amyloid Aggregators

Kenneth Osborne[1], Michael Bachmann[2], and Birgit Strodel[1]

[1] Institute of Complex Systems (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {k.osborne, b.strodel}@fz-juelich.de

[2] The University of Georgia, Center for Simulational Physics, Athens, GA 30602-2451, U.S.A
E-mail: bachmann@smsyslab.org

1   Introduction

A broad range of human diseases are known to arise as a consequence of protein aggregation and misfolding. A specific class of these so called amyloid proteins are known to generate fibrillar aggregates. Specifically these aggregates have a cross β-sheet structure, where the β-strands run perpendicular to the fiber's long axis, and the backbone hydrogen-bonds stabilize the sheets propagating along the direction of the fiber. Well known examples for diseases involving amyloid proteins include Alzheimer's disease, caused by aggregation of the Aβ peptide, and the transmissible Creutzfeldt-Jakob Disease, caused by aggregation of the human prion protein, PrP.

Although considerable progress has been made in recent years toward the elucidation of the structure and properties of amyloid fibrils, little is known about the structure and dynamics of the oligomers that are involved. The precise origin of pathogenicity in amyloid diseases remains elusive, although current evidence suggests that the soluble oligomeric precursors, rather than the fibrils themselves, are the cytotoxic species. Further studies that help to reveal the molecular mechanism of the multi-step process of amyloid aggregation are needed to find the missing link between amyloid fibrils and the disease to which they are connected.

The aim of the present work is to study the oligomerization of the peptide GNNQQNY. GNNQQNY is a polar heptapeptide from the N-terminal prion-determining region of the yeast prion protein Sup35. The atomic resolution crystal structure of GNNQQNY has recently been determined by Eisenberg and co-workers using x-ray microcrystallography[2]. There is strong evidence that the microcrystals formed by amyloidogenic peptides or proteins are closely related to their amyloid fibrils[3]. We aim to probe the self-assembled structure of GNNQQNY using peptideB, the coarse grained (CG) force field proposed by Bereau and Deserno[4]. We perform replica exchange molecular dynamics (REMD) simulations starting six monomeric GNNQQNY peptides in random starting positions, and analyze the results in terms of microcanonical and canonical quantities.

2   Methods

peptideB Force Field The GNNQQNY peptide (Gly-Asn-Asn-Gln-Gln-Asn-Tyr) was represented by the coarse grained (CG) force field peptideB[4]. The primary reason for using a coarse-grained force field was due to the time scales associated with peptide aggregation, which generally extends over a time scale much beyond a microsecond. Though molecular dynamics (MD) simulations of atomistic models have reached the microsecond time

scale, it is not possible to use standard MD simulations in explicit solvent to study aggregation processes extending over a time scale much beyond a microsecond. In the peptideB model the backbone is represented by three beads per residue, while only one bead per side chain is used. The side chain bead is the location of the C_β atom coining the name peptideB force field. The peptideB forcefield was parameterized to reproduce both local conformations and tertiary structures. The relatively high resolution of backbone beads allows for the force field to model physically relevant secondary structures, such as β-sheets, α-helicies, and random coils.

Replica Exchange Molecular Dynamics   The force field is used in conjunction with the Espresso simulation package described below. All simulations were run using REMD, where multiple MD runs of the same system (replicas) are run simultaneously at different temperatures. After a specified number of time steps, replicas at neighbouring temperatures can be exchanged, provided that a Metropolis criterion is satisfied. This procedure allows high-energy structures to be accepted for the replicas at higher temperature. The associated configurational changes then migrate to the replicas at lower temperatures when exchanged with each other. We performed 20 independent REMD simulations starting from six monomeric GNNQQNY peptides in random starting positions, each at 16 different temperatures ranging from approximately $189 - 366$ K. We used a cubic box with edge length 49.5 Å, giving a concentration of $C \sim 80$mM. We let the system relax for 10 μs before collecting statistics for analysis. Production runs lasted for 40 μs.

Analysis   For the identification of standard transitions and their characteristics, several impact parameters, such as the potential energy and two different orientational order-parameters, were collected during the REMD simulations and the heat capacity was calculated using the weighted histogram analysis method (WHAM)[5,7].

3   Results and Discussion

Low Energy   We found the minimum energy structure in every temperature thread of every simulation. Although the minimum energy structure does not give any thermodynamic information, is does provide a clue as to what peptideB might find to be the ground state of the system. As seen in Fig. 1, peptideB has successfully sampled a β-sheet aggregate, as is expected for GNNQQNY[2].

Heat Capacity   The heat capacity is a measure of the change in energy versus the change in temperature. As a system undergoes a major structural conformation change (phase transition), small changes in temperature produce proportionally larger changes in energy giving rise to a peak in the specific heat curve. In our work, we obtained the average specific heat as a function of temperature from a canonical WHAM analysis. Error in the specific heat was obtained by taking the standard deviation of the average specific heat of the system at the temperatures of the replicas over all 20 runs. As seen in the Fig. 2, the average heat capacity behaves in the manner expected for a structural transition, which transforms from the dissociated phase to the aggregated phase. The error bars tell us with certainty that the peak in the specific heat is not due to random fluctuations.
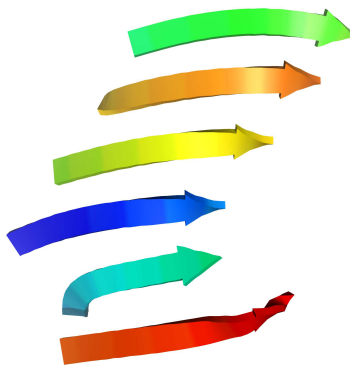
Figure 1. A minimum energy structure from one of the 20 GNNQQNY runs. This structure was sampled at $\sim$ 189 K. We see that although GNNQQNY forms β-sheets with parallel β-strands, the strands are not in-register.
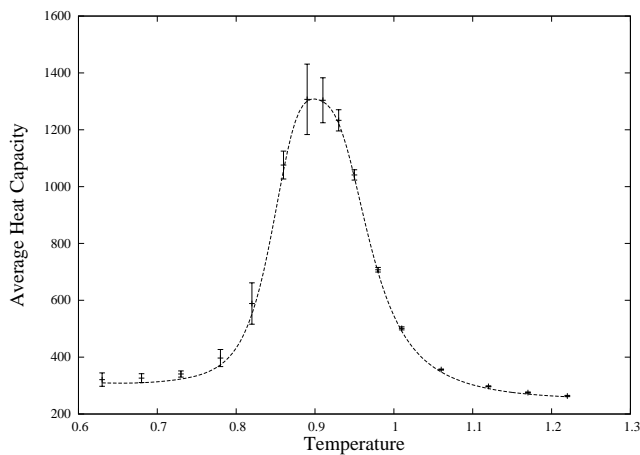

Figure 2. The heat capacity is plotted against temperature as obtained from our REMD simulations of six GN-NQQNY peptides in a box. Temperature in CG simulations is not as easy to define as temperature in atomistic simulations, but 1 temperature unit in the peptideB model is approximately equal to 300 K. Thus the peak at $\sim$ 0.9 temperature units indicates a phase transition at $\sim$ 267 K.

Orientational Order Parameters We also looked at the orientational order parameters $\overline{P_1}$ and $\overline{P_2}$, which are widely used for studying the properties of anisotropic fluids, and are
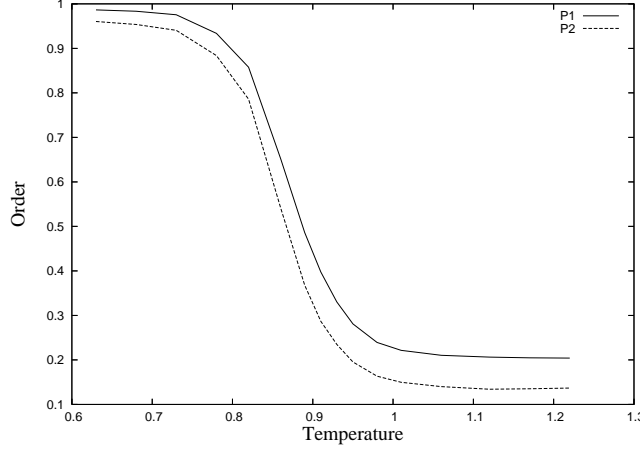
Figure 3. The average $\overline{P_1}$ and $\overline{P_2}$ as a function of temperature show how β-strand orientation evolves with temperature. Temperature in CG simulations is not strictly proportional to temperature outside of the simulation, but 1 Temperature Unit is approximately equal to 300 K.

defined as[6]:

$$\overline{P_1} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}} \right) \tag{1}$$

$$\overline{P_2} = \frac{1}{N} \sum_{i=1}^{N} \frac{3}{2} \left(\hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}} \right)^2 - \frac{1}{2} \tag{2}$$

Here, the director $\hat{\mathbf{d}}$ is a unit vector defining the preferred direction of alignment, $\hat{\mathbf{z}}_i$ is defined as unit vectors linking the peptides termini from the N to the C terminus, and N is the number of molecules in the simulation box, i.e., six peptides in this study. The director is defined as the eigenvector of the ordering matrix, that corresponds to the largest eigenvalue. $\overline{P_2}$ describes the orientational order of the system and discriminates between ordered and disordered conformations, i.e., $\overline{P_2}$ will be 1 if the monomers are arranged in either a parallel or antiparallel conformation and will vanish for the system in a fully isotropic state. The polar $\overline{P_1}$ describes the polarity of the system, i.e., how much the molecular vectors $\hat{\mathbf{z}}_i$ point in the same direction. It will be 1 if the monomers are arranged in a parallel conformation. $\overline{P_1}$ thus allows to discriminate between parallel and antiparallel mixed ordered aggregates. At low temperature, we see that $\overline{P_1}$ and $\overline{P_2}$ are high, whereas at high (~ 1) Temperature $\overline{P_1}$ and $\overline{P_2}$ are low (~ 0). At low temperature, we see that $\overline{P_1}$ and $\overline{P_2}$ are high, whereas at high (~ 1) Temperature $\overline{P_1}$ and $\overline{P_2}$ are low (~ 0).

With this knowledge, we can conclude that that peptideB has the capability of aggregating randomly structured monomers into ordered parallel beta-sheets.

Acknowledgments

154

References

1. D. J. Selkoe *Folding proteins in fatal ways*, Nature, **426**, 900–904, 2003.
2. R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Structure of the cross-β spine of amyloid-like fibrils*, Nature, **435**, 773–778, 2005.
3. M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, D. Eisenberg *Atomic structures of amyloid cross-beta; spines reveal varied steric zippers*, Nature, **447**, 453–457, 2007.
4. T. Bereau and M. Deserno, *Generic coarse-grained model for protein folding and aggregation*, J. Chem. Phys., **130**, 235106, 2009.
5. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *The weighted histogram analysis method for free-energy calculations on biomolecules. 1: The method* J. Comp. Chem., **13**, 1011–1021, 1992.
6. M. Cecchini, F. Rao, M. Seeber, A. Caflisch, *Replica exchange molecular dynamics simulations of amyloid peptide aggregation*, J. Chem. Phys., **121**, 10748–10756, 2004.
7. T. Bereau and R. H. Swendsen, *Optimized convergence for multiple histogram analysis.*, J. Comp. Phys., **228**, 6119, 2009.

155

Modeling Transmembrane Amyloid-β Structures: Aβ–Membrane Interactions

Chetan Poojari and Birgit Strodel

Institute of Complex Systems: Structural Biochemistry, Forschungszentrum Jülich
52425 Jülich, Germany
E-mail: {c.poojari, b.strodel}@fz-juelich.de

The etiology of Alzheimer's disease is considered to be linked to interactions between the amyloid-β (Aβ) peptide and neural cell membranes causing membrane disruption and increased ion conductance. The effects of $A\beta$ on lipid behavior have been characterized experimentally, but structural and causal details are lacking. We employ atomistic molecular dynamics (MD) simulations totaling over 4.5 microseconds in simulation time to investigate the behavior of $A\beta_{1-42}$ in zwitterionic and anionic lipid bilayers. We consider a transmembrane β-sheets (monomer and tetramer) and a helical structure preinserted in the membrane. In all MD simulations $A\beta_{1-42}$ remains embedded in the bilayer, with slow unfolding of the peptide monomer in the bilayer occurring in some cases. The N-terminal segment of the peptide outside the membrane strongly interacts with the lipid headgroups, leading to a disordering of the headgroup arrangement. We observe an increased stability for the β-sheet tetramer due to interpeptide interactions.

1   Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder associated with synaptic loss, abnormalities in functioning of neurons, neuronal cell death, and extracellular accumulation of senile plaques composed of the neurotoxic amyloid-β (Aβ) peptide. Aβ is derived from the amyloid precursor protein, a type 1 membrane integral glycoprotein through sequential cleavage by β- and γ-secretases producing 40 amino acids length (Aβ_{1-40}) and 42 amino acids length (Aβ_{42}) peptides. The more hydrophobic Aβ_{42} is the prevalent isoform seen in amyloid plaques and has a greater tendency to aggregate into fibrils and plaques. There is acceptable evidence suggesting that Aβ exerts its cytotoxic effect by strongly interacting with membranes of neurons and other cerebral cells. A potential pathway for Aβ toxicity lies in its ability to alter biophysical membrane properties[1–3]. Several studies have shown the ability of Aβ to form ion-channels in lipid bilayers causing an imbalance in calcium homeostasis. It was reported that permeabilization of membranes was caused by oligomers[4] but later Ambroggio et al.[5] showed the Aβ_{1-42} monomer to intercalate the membrane, too. There is sufficient experimental evidence demonstrating the effects of Aβ on membranes, but information about structural transformations of Aβ in membranes is still lacking. We use atomistic 500 ns MD simulations to examine the behavior of Aβ_{1-42} when preinserted into zwitterionic POPC and DPPC, and anionic POPG bilayers. The Aβ_{1-42} peptide remains within the bilayer throughout the simulations. Our results agree well with experimental findings by Mason et al.[6] where they reported that soluble Aβ_{1-40} peptide intercalates deep into the plasma membrane hydrocarbon core, whereas aggregated forms were seen interacting with the membrane bilayer at the headgroup-water interface. Though Aβ_{1-40} was the peptide under study[6], both Aβ_{1-40} and Aβ_{1-42} alloforms share common features in their ability to alter biophysical properties of membranes.
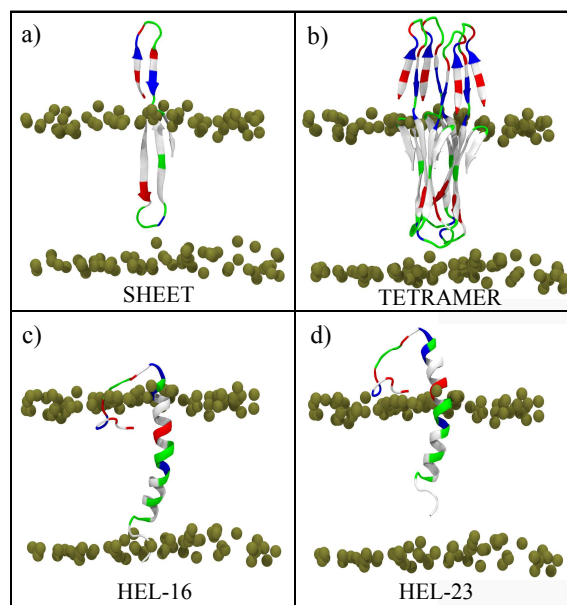
Figure 1. Initial structures for (a) monomeric β-sheet, (b) tetrameric monomeric β-sheet, (c) helix with K16 at the membrane-water interface (Hel-16), and (d) helix with D23 at the membrane-water interface (Hel-23). The peptide is rendered as a new cartoon and colored based on residue type, bilayer phosphorus atoms are shown as spheres, and bilayer tails and water molecules are not shown for clarity.

2 Methods

Starting structures The two initial $A\beta_{1-42}$ structures used in our simulations are a β-sheet and a helical conformation, respectively. The transmembrane β-sheet was obtained from a study for the $A\beta_{1-42}$ monomer and small oligomers with an implicit membrane model using a global optimization method[7]. The structure is comprised of an antiparallel β sheet with 2 turn regions within the membrane, and an N-terminal β-hairpin outside the membrane. The membrane-spanning β-sheet is inserted into the lipid bilayer such that the more hydrophobic residues (17-42) are located within the membrane, whereas the more hydrophilic residues (1-16) occupy the extracellular space. We study this transmembrane β-sheet as a monomer and a tetramer (Fig. 1(a) and (b)). The second starting structure used is a helical structure obtained from an NMR study of $A\beta_{1-40}$ in an apolar solvent (PDB 1BA4). We extended this 40 residue peptide by adding the two hydrophobic residues I41 and A42 to study the role of the extra two residues in peptide-lipid interactions. Previous studies have revealed an increased stability provided by I41 and A42 to the antiparallel β-sheet when compared to $A\beta_{1-40}$[8]. The helical structure was studied for two insertion depths: with K16 (Hel-16) and with D23 (Hel-23) at the membrane-water interface (Fig. 1(c) and (d)). All our simulations were carried out at the physiological charge of -3 for $A\beta_{1-42}$, with lysine and arginine residues protonated, aspartic and glutamic acid residues charged, and histidine residues modeled uncharged. The N- and C- terminals were capped to nullify the effect of terminal residues in peptide-lipid interactions.

Molecular dynamics MD simulations were performed using the GROMACS 4.0 package. The Aβ_{1-42} peptide was described using the GROMOS96 53A6 force field and the lipid models with modified Berger force field parameters for use with the GROMOS96 53A6 force field[9]. Initial coordinates of 128 lipids for POPC, DPPC and POPG bilayers, equilibrated with water for 40 ns were obtained from Kukol's work on lipid models[9]. The Aβ_{1-42} peptide was inserted into the pre-equilibrated lipid membrane using the Inflategro script[10], the whole system then solvated with SPC water molecules, Na$^+$ counterions were added to balance the peptide charge, and 0.1M NaCl salt added to bring the system to the near physiological salt concentration. The simulations were carried out in a 6.5 × 6.5 × 9.5 nm box. An initial equilibration under isothermal-isochoric (NVT) conditions was performed for 100 ps during which the protein heavy atoms and phosphorous atoms of the lipid headgroups were restrained. Here, the V-rescale thermostat was used to regulate the temperature of the Aβ_{1-42} peptide, lipids and solvent/ions separately at 298K for the POPC and POPG simulations and at 325K for the DPPC simulations. After this, the systems were equilibrated for 30 ns under isothermal-isobaric (NPT) conditions using the Nose-Hoover thermostat along with semiisotropic Parrinello-Rahman pressure coupling. The bilayer normal z-direction and xy plane were coupled separately maintaining a constant pressure of 1 bar independently in all directions. Long-range electrostatics were calculated using the Particle Mesh-Ewald method in connection with periodic boundary conditions. Van der Waals and Coulombic interaction cutoffs were set to 1.2 nm and the LINCS algorithm was used to constrain all bond lengths. Following equilibration, production MD runs were performed for 500 ns for each system. Here the parameter settings were similar to the NPT equilibration step but with all restraints being removed. The time step for integration was 2 fs with coordinates and velocities saved every 20 ps for analysis.

3   Results and Discussion

The final structures after the 500 ns MD simulations of monomeric and tetrameric Aβ_{1-42} are shown in Fig. 2. In all simulations the peptide remained embedded in the lipid bilayer and structural changes occur to a varying degree mainly depending on the lipid type. An increased stability is observed for the Aβ_{1-42} tetramer preinserted in a POPC bilayer.

Aβ_{1-42} in POPC bilayer Our results for the POPC bilayer are in accord with experimental results of Ambroggio et al.[5] where the Aβ_{1-42} peptide remained well embedded in the lipid environment composed of POPC or POPC/SM/Chol altering cohesion between the membrane components and membrane permeability. The stable β-sheet structure seen in our simulations is supported by experimental work[11] where it was shown that Aβ_{1-40} is present as a β-sheet in a POPC bilayer. A further study reported β-sheet content in the membrane hydrophobic core of Aβ_{1-42} incorporated into POPC/POPS mixed bilayers, which destabilizes the membrane, thereby increasing its permeability properties[12]. Similar to the β-sheet, the helical peptide did also not exit the bilayer. The Hel-16 system moves upwards so that residues E22/D23 reach the membrane-water interface. Both the sheet and helix structures were stable within the membrane. A further increased stability is observed for the β-sheet tetramer both within and outside the membrane. Within the membrane the β-sheet is stabilized due to intermolecular salt bridges between D23 and K28, and side chain interactions formed between interpeptide hydrophobic residues. Outside the
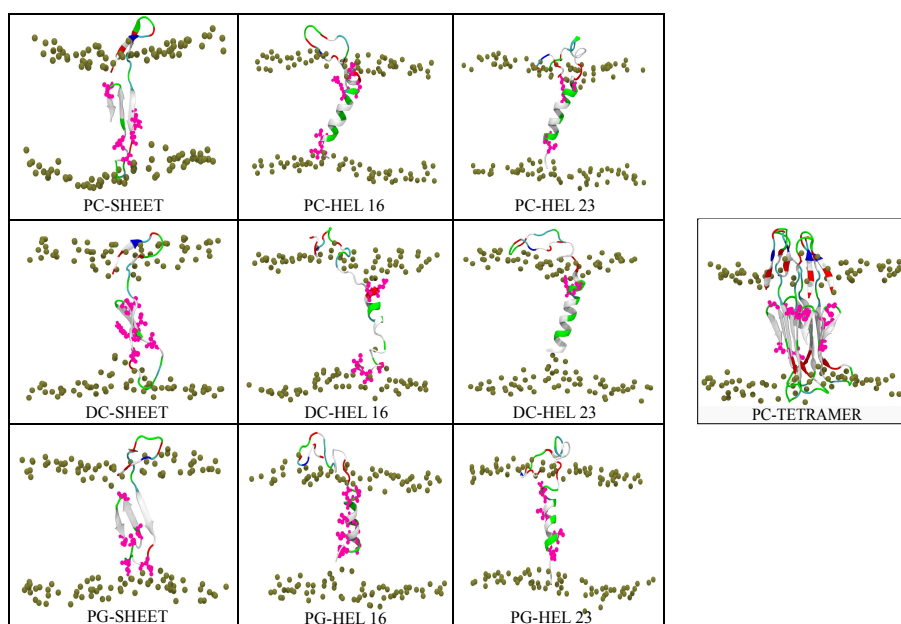
Figure 2. Final structure from the 500 ns MD simulations with (left) monomeric Aβ_{1-42} in a POPC (PC), DPPC (DC) and POPG (PG) bilayer; (right) tetrameric Aβ_{1-42} in a POPC bilayer. See Fig. 1 for explanation of the graphical representation. The peptide residues rendered as pink spheres are involved in H-bond formation with the water molecules entering the bilayer.

membrane the N-terminal β-hairpins form stable interactions with each other instead of with the bilayer surface. Similar to our findings Lal and co-workers showed that Aβ_{1-42} forms stable tetramers when incorporated within the membrane[13].

Aβ_{1-42} in DPPC bilayer Simulations with the DPPC bilayer show in general a larger loss of secondary structure content for both sheet and helical conformations within the membrane core compared to the POPC simulations. The loss of β content for the preinserted β-sheet facilitates excessive entry of water molecules into the hydrophobic core of the bilayer. The helical peptide inserted at K16 looses its α-helix secondary structure to a considerable extent within the membrane, whereas the peptide inserted at D23 adopts a more stable α-helix within the membrane. The Aβ_{1-42} residues outside the membrane interact strongly with the lipid headgroups, leading to the N-terminal section laying parallel to the bilayer surface with loss of the helical structure. Our findings for the DPPC bilayer do not completely correspond to the results of MD simulations performed by Xu et al.[14], where they carried out simulations with Aβ_{1-40} (PDB 1BA4) preinserted with K28 at the membrane-water interface. They observed Aβ_{1-40} exiting the hydrophobic core and associating with the bilayer surface. They also reported that Aβ_{1-40} maintains an α-helix as secondary structure even at the membrane-water interface. Such difference in results are probably due to the different insertion depths of the peptide, different force fields employed in the studies, and different methods used to insert the peptide into the lipid bilayer. Simulations carried out by Lemkul et al.[15] with the same Aβ_{1-40} structure in a DPPC bilayer

reported the peptide to remain partially embedded in the bilayer when inserted with K28 at the membrane water interface and observed complete loss of α-helicity within the first 10 ns. When Aβ_{1-40} peptide was inserted with K16 at the membrane water interface, the peptide remained embedded in the bilayer and retained α-helicity in the central segment[15]. These findings agree with our results for Aβ_{1-42} with K16 is at membrane-water interface. The presence of peptide within the central hydrophobic core of the membrane is supported by experimental findings of Mason et al.[6]

Aβ_{1-42} in POPG bilayer The simulation results for the POPG bilayer are similar to our results with the POPC bilayer, which shows that once the peptide has entered the membrane the peptide behavior in the hydrophobic core is dominated by the amount of hydrophobic mismatch between the hydrophobic length of the Aβ_{1-42} peptide and the hydrophobic thickness of the membrane it spans. The head-group type is of less importance for the stability of the peptide in the membrane. Since the hydrophobic mismatch between Aβ_{1-42} and POPC and POPG, respectively is smaller than for DPPC, transmembrane Aβ_{1-42} is more stable in the two former lipid bilayers. The Hel-16 peptide moves upwards so that E22/D23 reach the bilayer-water interface as is also observed in the POPC bilayer. The β-sheet and helical structures of Aβ_{1-42} in the POPG bilayer allowing excessive entry of water molecules into the hydrophobic core of the bilayer.

Acknowledgments

References

1. E Terzi, G Holzemann, and J Seelig, *Interaction of Alzheimer beta-amyloid peptide(1-40) with lipid membranes*, Biochemistry, **36**, 14845–14852, 1997.
2. A Buchsteiner, T Hauss, S Dante, and NA Dencher, *Alzheimer's disease amyloid-beta peptide analogue alters the ps-dynamics of phospholipid membranes*, BBA-Biomembranes, **1798**, 1969–1976, 2010.
3. GP Eckert, WG Wood, and WE Mueller, *Lipid Membranes and beta-Amyloid: A Harmful Connection*, Curr. Protein Pept. Sci., **11**, 319–325, 2010.
4. R Kayed, Y Sokolov, B Edmonds, TM McIntire, SC Milton, JE Hall, and CG Glabe, *Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases*, J. Biol. Chem., **279**, 46363–46366, 2004.
5. EE Ambroggio, DH Kim, F Separovic, CJ Barrow, CJ Barrow, KJ Barnham, LA Bagatolli, and GD Fidelio, *Surface behavior and lipid interaction of Alzheimer beta-amyloid peptide 1-42: A membrane-disrupting peptide*, Biophys. J., **88**, 2706–2713, 2005.
6. RP Mason, RF Jacob, MF Walter, PE Mason, NA Avdulov, SV Chochina, U Igbavboa, and WG Wood, *Distribution and fluidizing action of soluble and aggregated amyloid beta-peptide in rat synaptic plasma membranes*, J. Biol. Chem., **274**, 18801–18807, 1999.

7. B Strodel, JWL Lee, CS Whittleston, and DJ Wales, *Transmembrane Structures for Alzheimer's A beta(1-42) Oligomers*, J. Am. Chem. Soc., **132**, no. 38, 13300–13312, 2010.
8. DL Mobley, DL Cox, RRP Singh, MW Maddox, and ML Longo, *Modeling amyloid beta-peptide insertion into lipid bilayers*, Biophys. J., **86**, 3585–3597, 2004.
9. A Kukol, *Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins*, J. Chem. Theor. Comput., **5**, 615–626, 2009.
10. C Kandt, WL Ash, and DP Tieleman, *Setting up and running molecular dynamics simulations of membrane proteins*, Methods, **41**, 475–488, 2007.
11. MRR de Planque, V Raussens, SA Contera, DTS Rijkers, RMJ Liskamp, J-M Ruysschaert, JF Ryan, F Separovic, and A Watts, *beta-sheet structured beta-amyloid(1-40) perturbs phosphatidylcholine model membranes*, J. Mol. Biol., **368**, 982–997, 2007.
12. TL Lau, EE Ambroggio, DJ Tew, R Cappai, CL Masters, GD Fidelio, KJ Barnham, and F Separovic, *Amyloid-beta peptide disruption of lipid membranes and the effect of metal ions*, J. Mol. Biol., **356**, 759–770, 2006.
13. H Lin, R Bhatia, and R Lal, *Amyloid beta protein forms ion channels: implications for Alzheimer's disease pathophysiology*, FASEB JOURNAL, **15**, 2433–2444, 2001.
14. YC Xu, JJ Shen, XM Luo, WL Zhu, KX Chen, JP Ma, and HL Jiang, *Conformational transition of amyloid beta-peptide*, Proc. Natl. Acad. Sci. USA, **102**, 5403–5407, 2005.
15. JA Lemkul and DR Bevan, *A comparative molecular dynamics analysis of the amyloid beta-peptide in a lipid bilayer*, Arch. Biochem. Biophys., **470**, 54–63, 2008.

Dynamical Conformation of Biantennary Complex-type N-glycan in Water Revealed by Using Replica-exchange Molecular Dynamics Simulations

Suyong Re[1], Naoyuki Miyashita[2], Yoshiki Yamaguchi[3], and Yuji Sugita[1]

[1] Theoretical Biochemistry Laboratory, RIKEN Advanced Science Institute
2-1 Hirosawa, Wako, Saitama 351-0198, Japan
E-mail: {suyongre, sugita}@riken.jp

[2] RIKEN Advanced Institute for Computational Science, Hyogo 650-0047, Japan
E-mail: miya@riken.jp

[3] Structural Glycobiology Team, RIKEN Advanced Science Institute
2-1 Hirosawa, Wako, Saitama 351-0198, Japan
E-mail: yyoshiki@riken.jp

Here we report replica-exchange molecular dynamics (REMD) simulation of a biantennary complex-type N-glycan in water, fully characterizing the dynamical conformation. The effect of the common N-glycan modification, introduction of the bisecting GlcNAc, is also examined. Five distinct conformers were found for the N-glycan. The conformational variety was significantly reduced upon the introduction of bisecting GlcNAc, leaving only two forms as major conformers. The results provide an insight into the mechanism of "conformation selection" for the specific binding to the target protein.

1 Introduction

Dynamical aspect of oligosaccharide conformation is a clue to understand glycan-protein recognitions that regulate many biological processes. Multiple conformations of glycan and their modification for instance by chemical substitutions are important for the specific binding to the target protein. On the other hand, the flexible motion through many stereoisomers makes the characterization of global conformation extremely challenging. In general, glycans are not well resolved in X-ray crystal structures. Nuclear magnetic resonance (NMR) spectroscopy-based analysis is used more frequently, providing the dynamical information of glycosidic linkage. Molecular dynamics (MD) simulation is a powerful alternative. It not only complements experimental data but also reveals the dynamical motion of tertiary structures. Nevertheless, the simulation is limited to small glycans in water, because the conventional MD simulation hardly samples all possible conformational space.

Our aim is to characterize the global conformation of glycan that is important for the selective binding to proteins. To that end, we applied replica-exchange MD (REMD) method[1] to reveal the energy landscape of conformational space for this extremely flexible molecule.

2 Computational Method

We performed REMD simulations of a biantennary complex-type N-glycan with (Bi9) and without (BiB10) the chemical modification, bisecting GlcNAc (Fig. 1). Glycam06 force
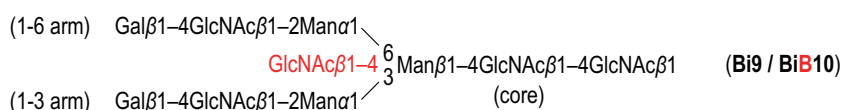
163

(1-6 arm)   Galβ1–4GlcNAcβ1–2Manα1

GlcNAcβ1–4 $^6_3$ Manβ1–4GlcNAcβ1–4GlcNAcβ1   (**Bi9 / BiB10**)
(core)

(1-3 arm)   Galβ1–4GlcNAcβ1–2Manα1

Figure 1. Sequence of biantennary N-glycan Bi9 and BiB10.

field parameter[2] for oligosaccharide and TIP3P model for water molecules were used. 64 replicas were used with the temperature range from 300 K to 500 K. The replica exchange was tried every 1ps. REMD simulations were run for 3.3-μs in total. The average acceptance ratios of replica exchange was 49 %. In Fig. 2 we show the time series of replica exchange at the lowest temperature (T = 300 K) as well as temperature exchange. We observed random walks in both the 'replica space' and the 'temperature space.'
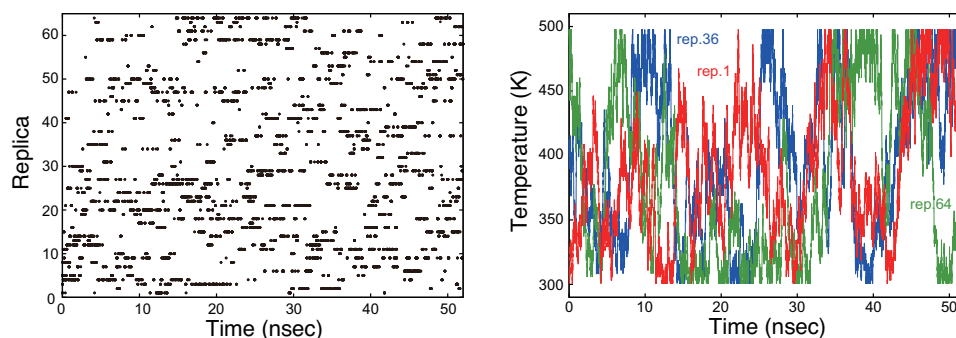
Figure 2. Time series of replica exchange at T = 300 K (left) and temperature exchange (right).

3   Results

3.1   Glycosidic Linkage Conformation

Fig. 3A shows the free-energy maps of the orientation of Manα1-6Manβ1 linkage for Bi9. The linkage adopts five distinct conformers which differ in dihedral angles Ψ (C_1-O-C_6'-C_5') and (O-C_6'-C_5'-C_4'). Note that 50 ns conventional MD simulation completely missed some of important conformers (Fig. 3B). The dihedral angle Φ (C_5-C_1-O-C_6') is rather uniform, $\sim$70°, in consistent with the *exo-anomeric effect*. Ψ angle has three minima around 60° (**a**), 96° (**b**), and -174° (**c**). First two conformers are energetically favored. These conformers preferentially adopt *gauche-gauche* orientation (= 42° (**a**), 48° (**b**)), although *gauche-trans* orientation is also possible (ω = 168° (**a'**)). Three conformers (**a**, **b**, and **a'**) dominate over the other two (**c** and **c'**). The bisecting GlcNAc shifts the preference of Ψ angle from 60° (**a**) and 90° (**b**) for Bi9 to 90° (**a**) and 180° (**c**) for BiB10 (Fig. 3C). Two conformers (**a** and **a'**) are significantly less populated, whereas the population of conformer **c** increases.
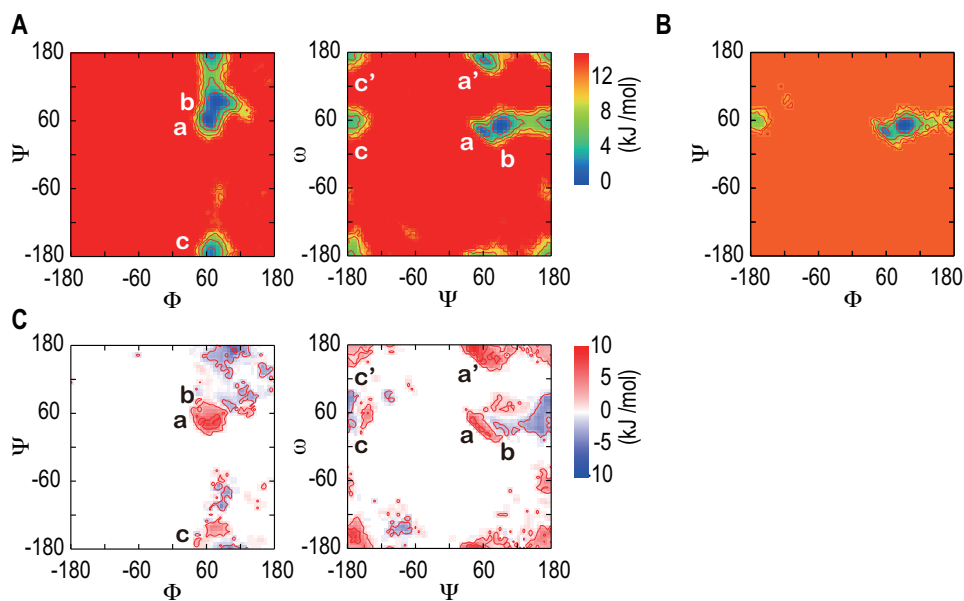
Figure 3. (A) Φ, Ψ and Ψ, ω free-energy maps for Bi9. (B) Ψ, ω free-energy map for Bi9 from 50 ns conventional MD. (C) Difference free-energy maps (BiB10 - Bi9).

3.2 Global Conformation of Bi9 and BiB10

Fig. 4 shows the representative conformers and their populations obtained from the clustering analysis for Bi9 (A) and BiB10 (B). The global conformation nicely correlates to the local Manα1-6Manβ1 linkage orientation. Three low-energy linkage conformers (**a**, **b**, and **a'**) lead to Backfold type, whereas the less favored conformers (**c** and **c'**) lead to

(a') "Tight Backfold" (13%) (b) "Backfold" (41%) (a) "Half Backfold" (17%) (b) "Backfold" (59%)

(c) "Extend-a" (10%) (c') "Extend-b" (13%) (c) "Extend-a" (29%)
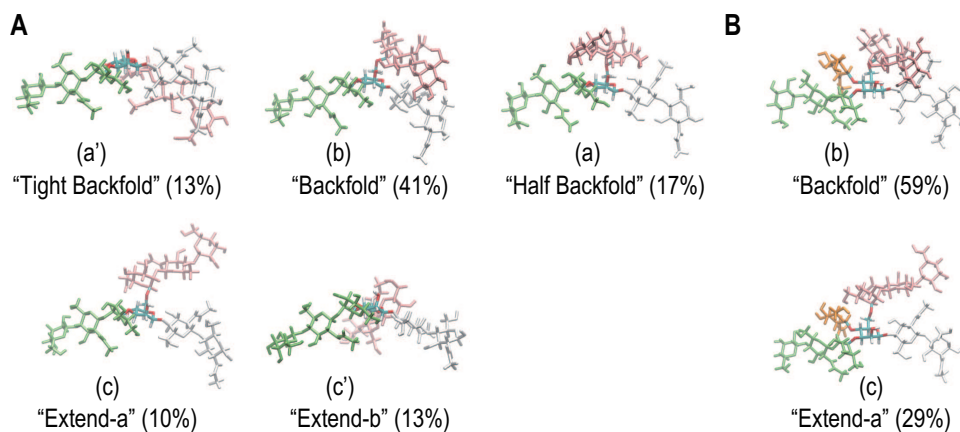
Figure 4. Representative conformers and populations obtained from the clustering analysis.

Extend type. The bisecting GlcNAc suppresses the conformational variety, leaving only Backfold and Extend-a forms as major conformers.

N-glycan modification is known to alter several functional activity[3]. Change in the binding affinity to the target protein is important. Early NMR-based-analyses of model oligosaccharides[4,5] and recent MD simulation[6] support the idea that chemical modifications induce unique conformation suitable for the target protein. Our results suggest that chemical modifications may not induce any unique conformation but limit the conformational flexibility (select only suitable conformation) for the target protein.

4   Concluding Remarks

We performed REMD simulation of a biantennary complex-type N-glycan in solution. We have fully characterized the dynamical conformation of the N-glycan and that with the bisecting GlcNAc. We have shown that the bisecting GlcNAc does not induce unique conformation but limits the conformational flexibility. The results provide an insight into the mechanism of "conformation selection" for the specific binding to the target protein.

Acknowledgments

References

1. Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. **314**, 141–151, 1999.
2. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, R. J. Woods, *GLYCAM06: A generalizable biomolecular force field. Carbohydrates*, J. Comput. Chem. **29**, 622–655, 2008.
3. M. Takahasi, Y. Kuroki, K. Ohtsubo, and N. Taniguchi, *Core fucose and bisecting GlcNAc, the direct modiers of the N-glycan core: their functions and target proteins*, Carbohydrate. Res. **344**, 1387–1390, 2009.
4. S. W. Homans, R. A. Dwek, and T. W. Rademacher, *Tertiary structure in N-linked oligosaccharides*, Biochemistry. **26**, 6553–6560, 1987.
5. S. W. Homans, R. A. Dwek, and T. W. Rademacher, *Solution conformations of N-linked oligosaccharides*, Biochemistry. **26**, 6571–6578, 1987.
6. S. André, T. Kožár, S. Kojima, C. Unverzagt, and H-J. Gabius, *From structural to functional glycomics: core substitutions as molecular switches for shape and lectin affinity of N-glycan*, Biol. Chem. **390**, 557–565, 2009.

Neurodegenerative Diseases: A Molecular View

Parkinson disease: **Valeria Losasso**[1], **Domenica Dibenedetto**[1]
Huntington disease: **Giulia Rossetti**[1]
Prion disease: **Giulia Rossetti**[1], **Xiaojing Cong**[2]
PI: **Paolo Carloni**[1]

[1] German Research School for Simulation Sciences, 52425 Jülich, Germany
E-mail: p.carloni@grs-sim.de

[2] International School for Advanced Studies (SISSA), 34136 Trieste, Italy

We will present computational approaches to address three neurodegenerative disorders: Parkinson disease (PD), Huntington Disease (HD) and Prion disease (or transmissible spongiform encephalopathies, TSEs). PD is characterized by a selective degeneration of specific subsets dopamine neurons. The formation of cytoplasmic aggregates called Lewy bodies, mainly constituted by α-synuclein is an hallmark of the disease. HD is caused by an expanded CAG repeat in the huntingtin gene encoding for polyglutamine tract in the huntingtin protein (Htt). Neuronal intra-nuclear and intra-cytoplasmic inclusions rich in polyglutamine are pathological hallmarks of the disease. TSEs are characterized by the conformational change of a membrane bound protein, the cellular PrP^C, into a disease associated, fibril-forming isoform, PrP^{Sc}.

1 Introduction

An increasing family of neurodegenerative disorders is associated with aggregation of misfolded polypeptide chains, which are toxic to the cell. It is essential to understand why and how endogenous proteins may adopt a non-native fold and how to develop new drugs with improved efficacy.

PD. We have studied two different aspects of α-synuclein (AS), namely: (1) the structural role of compensatory replacements in mice AS, where a point mutation linked to human familial PD is physiologically present in the wild-type sequence; (2) the modulation of AS aggregation by dopamine (DOP) analogs and metabolism products. (1) Three point mutations in the AS gene, A30P, E46K and A53T have been identified in familial PD cases[1]. In non-Primate mammals, A is replaced by T, that is the familial mutation in PD. In mice, 53T is accompanied by six other substitutions (S87N, L100M, N103G, A107Y, D121G and N122S). To explore the structural role of the compensatory aminoacid changes in mice, we performed a comparative sub-microsecond molecular dynamics (MD) simulation study of wild-type human AS (huAS(wt)), A53T-mutated human AS (huAS(A53T)) and mouse AS (mAS(wt)). (2) AS has been shown to regulate dopamine metabolism at multiple levels including its synthesis, uptake and release. AS fibrillization and DOP metabolism are likely to be linked to PD pathogenesis[2]. DOP, as well as several products derived from its oxidation, bind to the AS C-terminal region comprising the YEMPS residues (125-129). These molecules inhibit the conversion of AS to mature fibrils, promoting instead, accumulation of innocuous oligomeric forms[3]. We applied an integrated computational and experimental approach to find alternative ligands that might modulate the AS fibrillization.

HD. The first exon of the huntingtin protein (Htt Exon 1) contains the polyglutamine (polyQ) region. It is implicated in Huntington disease[4,5]. We have investigated by computational methods two relevant structural aspects of Htt Exon 1. (1) The length of the polyQ region in Exon 1 is inversely correlated to the age of onset of symptoms[6]. The similarity between the Q side chain and the amino acids backbone attributes a crucial role to the formation of intramolecular hydrogen bonds (HB) in polyQ aggregates[7]. We have studied the effect of cooperativity on the polyQ expansion for understanding if HB content plays a major role in the structural stability of polyQ systems. (2) The first seventeen amino acids of the N-term region (N17) modulate aggregation and toxicity of Htt Exon 1[8]. Aggregation might arise by a variety of mechanisms[8], the understanding of which greatly benefits from structural information. However, neither the secondary structure content nor the tertiary structure for N17 is known. We describe the thermodynamics and the kinetics of N17 in aqueous solution at room temperature.

TSEs. TSEs arise with the post-translational conversion of the ubiquitous cellular form of the prion protein, PrP^C, into its pathogenic form, PrP^{Sc}, without any detectable covalent modifications[9]. A relationship between point mutations and TSEs has been firmly established[10]. Indeed, in the presence of specific disease-linked mutations, the PrP^C to PrP^{Sc} conversion appears to occur spontaneously[11]. The mutations accelerate the kinetics of the misfolding process relative to WT, possibly mostly because of a destabilization of the native structure and/or an increasing in the stability of the partially folded intermediate species[11,12]. Hence, we have investigated the intrinsic instability of PrP^C fold mutants as key for TSEs research.
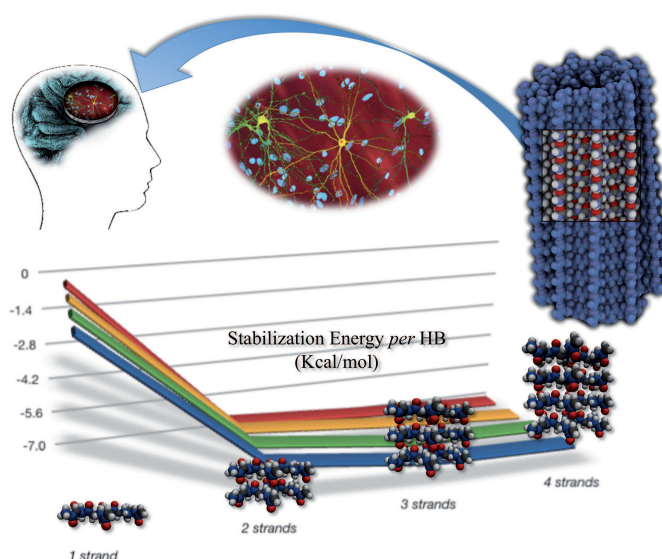

Figure 1. *Cooperative Effect of PolyQ.* Stabilization Energy *per* HB: the strength of the HBs between layers increases nonlinearly with the number of strands, according to our calculations.
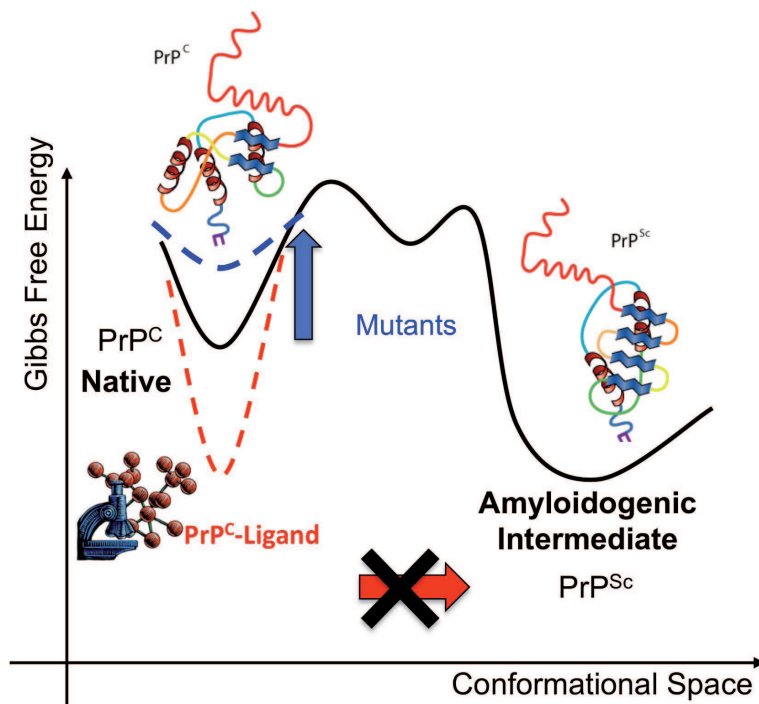
Figure 2. PrP^C *to* PrP^{Sc} *conversion.* Qualitative scheme illustrating the Gibbs free energy change in the interconversion, as proposed by Prusiner et al.[9]. The interconversion is thought to entail intermediate partially folded conformers. The mutation could destabilize PrP^C (hence increasing the population of intermediates), whilst small ligands may stabilize the PrP^C conformation.

2   Methods

PD. (1) To study the role of the compensatory aminoacid substitutions from human to the mouse protein, we performed classical MD simulations with NAMD[13], which were extracted from a NMR ensemble[14]. (2) The five selected ligands were docked onto the 6 AS conformers using AutoDock[15].

HD. We use classical[16] and *ab-initio*[17] MD for studying polyQ properties. Bias Exchange metadinamics[18] and modeler[19] for N17 features.

TSEs. We apply molecular dynamics (MD) simulations using a previously established MD protocol[20].

3 Results

PD. (1) Our calculations on huAS(wt) and huAS(A53T) reproduced the experimental structural information[21]. We then took advantage of the same computational protocol to identify the structure and conformational fluctuations of mAS(wt). Structure and plasticity changes in mAS(wt) protein restore structural features not too dissimilar from those of huAS(wt). Future experiments in vivo and in vitro should address the role of S87N substitution as an intramolecular suppressor of A53T toxicity. (2) Three of the screened ligands (6-aminoindole, 5-hydroxyindole, 2-amino-4-tert-butylphenol) were shown by MD simulation to interact significantly with AS (and in particular with the 125YEMPS129 region), although to a lesser extent than most dopamine derivatives. In agreement with these predictions, in vitro assays in combination with high-resolution microscopy revealed the inhibitory effect of these ligands on AS aggregation. The consistency between both approaches highlights the importance of a combined in silico and in vitro approach that could be used in predicting and developing new drugs and therapeutic strategies for PD.

HD. We have further corroborated the hypothesis[7] that the HB content plays a major role in the structural stability of polyQ systems. Indeed, we find that PolyQs tracts always tried to establish the higher number of HBs between both side-chain and backbone atoms, independently of the β-sheet content and the number of Qs in each structure[22]. Furthermore, quantum mechanical calculations point to the fact that, as expected, the HB network of polyQ β-sheets is associated with a cooperative effect (CE) other than that of the amino acids[23]. The presence of such extra-stabilization may explain polyQ aggregation propensity. Next, we have investigated structure and binding to a cellular partner of Htt Exon 1. First, we have predicted the conformational properties of the peptide in solution. The results are in accord with experiments[24]. The detailed structural characterization of N17 alone may helps to investigate N17 interactions to its cellular partners.

TSEs. MD calculations, carried out on all of the pathogenic mutants of Pr^C globular domain show that all the variants feature common traits, independently from position and chemical nature of the mutations. These include: (i) the disruption of a specific salt bridge network present in HuPrP(WT) in all the disease linked mutants (DLMs), with consequent increase in flexibility of α_3 in most of them. (ii) The loss or the weakening of hydrophobic interactions present in HuPrP(WT) in all the DLMs. (iii) An increased solvent exposure of Y169 in $\beta_2 - \alpha_2$ loop relative to HuPrP(WT) in most DLMs. From the biological point of view, the altered conformation observed in HuPrP mutants might cause a different affinity for cellular membranes and, consequently, an aberrant localization of PrP in cellular compartments, favoring formation of altered endoplasmatic reticulum topologies[25]. Independent evidence derived from cell culture, expressing some of the DLMs studied here, showed that these mutations affect folding and maturation of Pr^C in the secretory pathway of neuronal cells[26]. The structure-function relationships suggested by this work may contribute to understand the molecular basis of the generation of PrP^{Sc} in inherited prion diseases.

Acknowledgments

References

1. JM Shulman, PL De Jager, and MB Feany, *Parkinson's disease: genetics and pathogenesis*, Annu Rev Pathol, **6**, 193–222, 2011.

2. TM Fountaine and R Wade-Martins, *RNA interference-mediated knockdown of alphasynuclein protects human dopaminergic neuroblastoma cells from MPP(+) toxicity and reduces dopamine transport*, J Neurosci Res, **85**, no. 2, 351–63, 2007.

3. FE Herrera, A Chesi, KE Paleologou, A Schmid, A Munoz, M Vendruscolo, S Gustincich, HA Lashuel, and P Carloni, *Inhibition of alpha-synuclein fibrillization by dopamine is mediated by interactions with five C-terminal residues and with E83 in the NAC region*, PLoS ONE, **3**, no. 10, e3394, 2008.

4. SW Davies, M Turmaine, BA Cozens, M DiFiglia, AH Sharp, CA Ross, E Scherzinger, EE Wanker, L Mangiarini, and GP Bates, *Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation*, Cell, **90**, no. 3, 537–548, 1997.

5. L Mangiarini, K Sathasivam, M Seller, B Cozens, A Harper, C Hetherington, M Lawton, Y Trottier, H Lehrach, SW Davies, and GP Bates, *Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice*, Cell, **87**, no. 3, 493–506, 1996.

6. C Ross, *When more is less: Pathogenesis of glutamine repeat neurodegenerative diseases*, Neuron, **15**, no. 3, 493–496, 1995.

7. M Perutz, *Polar Zippers - Their Role in Human-Disease*, Protein Sci, **3**, no. 10, 1629–1637, 1994.

8. R Truant, RS Atwal, C Desmond, L Munsie, and T Tran, *Huntington's disease: revisiting the aggregation hypothesis in polyglutamine neurodegenerative diseases*, Febs J, **275**, no. 17, 4252–4262, 2008.

9. S Prusiner, *Novel Proteinaceous Infectious Particles Cause Scrapie*, Science (New York, NY), **216**, 136–144, 1982.

10. MW van der Kamp and V Daggett, *The consequences of pathogenic mutations to the human prion protein*, Protein Eng Des Sel, **22**, no. 8, 461–468, 2009.

11. AC Apetri and K Surewicz, *The effect of disease-associated mutations on the folding pathway of human prion protein*, Journal of Biological Chemistry, **279**, no. 17, 18008–18014, 2004.

12. K Surewicz, EM Jones, and AC Apetri, *The emerging principles of mammalian prion propagation and transmissibility barriers: Insight from studies in vitro*, Accounts Chem Res, **39**, no. 9, 654–662, 2006.

13. JC Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, RD Skeel, L Kalé, and K Schulten, *Scalable molecular dynamics with NAMD*, J Comput Chem, **26**, no. 16, 1781–802, 2005.

14. MM Dedmon, K Lindorff-Larsen, J Christodoulou, M Vendruscolo, and CM Dobson, *Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations*, J Am Chem Soc, **127**, no. 2, 476–7, 2005.

15. G Morris, D Goodsell, R Halliday, R Huey, W Hart, R Belew, and A Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*, Journal Of Computational Chemistry, **19**, no. 14, 1639–1662, 1998.

16. B Hess, C Kutzner, D van der Spoel, and E Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput., **4**, no. 3, 435–447, 2008.

17. CPMD, `http://www.cpmd.org/`.

18. S Piana and A Laio, *A bias-exchange approach to protein folding*, The journal of physical chemistry B, **111**, no. 17, 4553–9, 2007.

19. R Sánchez and A Sali, *Evaluation of comparative protein structure modeling by MODELLER-3*, Proteins, **Suppl 1**, 50–8, 1997.

20. G Rossetti, G Giachin, G Legname, and P Carloni, *Structural facets of disease-linked human prion protein mutants: A molecular dynamic study*, Proteins: Structure, Function, and Bioinformatics, **78**, no. 16, 3270–3280, 2010.

21. V Losasso, A Pietropaolo, C Zannoni, S Gustincich, and P Carloni, *Structural Role of Compensatory Amino Acid Replacements in the α-Synuclein Protein*, Biochemistry, pp. 6994–7001, 2011.

22. G Rossetti, A Magistrato, A Pastore, F Persichetti, and P Carloni, *Structural properties of polyglutamine aggregates investigated via molecular dynamics simulations.*, Journal Of Physical Chemistry B, **112**, no. 51, 16843–16850, 2008.

23. G Rossetti, A Magistrato, A Pastore, and P Carloni, *Hydrogen Bonding Cooperativity in polyQ β-Sheets from First Principle Calculations*, Journal Of Chemical Theory And Computation, **6**, no. 6, 1777–1782, 2010.

24. AK Thakur, M Jayaraman, R Mishra, M Thakur, VM Chellgren, I-JL Byeon, DH Anjum, R Kodali, TP Creamer, JF Conway, AM Gronenborn, and R Wetzel, *Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism*, Nat Struct Mol Biol, **16**, no. 4, 380–389, 2009.

25. J Heske, U Heller, KF Winklhofer, and J Tatzelt, *The C-terminal globular domain of the prion protein is necessary and sufficient for import into the endoplasmic reticulum*, Journal of Biological Chemistry, **279**, no. 7, 5435–43, 2004.

26. A Ashok and RS Hegde, *Selective Processing and Metabolism of Disease-Causing Mutant Prion Proteins*, PLoS Pathog, **5**, no. 6, e1000479, 2009.

Molecular Dynamics Simulations of Ribosome-oxazolidinone Complexes Reveal Structural Aspects for Antibiotics Design

Jagmohan Saini, Simone Fulle, and Holger Gohlke

Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University
Universitätsstr. 1, 40225 Düsseldorf, Germany
E-mail: {*saini, gohlke*}*@uni-duesseldorf.de*

We present an investigation of the determinants of binding of oxazolidinone antibiotics, in particular linezolid, its derivative radezolid, and the structurally related oral anticoagulant drug rivaroxaban in complex with the large ribosomal subunit from *Haloarcula marismortui* using molecular dynamics simulations. Our results are in agreement with available experimental data that show radezolid as the most potent inhibitor compared to linezolid and rivaroxaban. Furthermore, the structural and dynamical insights obtained from this study will provide ability to design improved antibiotics in the future.

1   Introduction

The ribosome is a large ribonucleoprotein complex that carries out protein synthesis in all kingdoms of life. It is composed of a large and a small subunit, which are denoted as 50S and 30S in bacteria. Peptide synthesis can be inhibited by antibiotics that bind to three functionally different sites on the ribosomal structure, namely the decoding site, the peptidyl-transferase center (PTC), and the protein exit tunnel[1]. High resolution crystal structures of large and small ribosomal subunits in complex with antibiotics[2,3] have revolutionized our understanding of their binding sites, binding modes, and mechanisms of action. Such information provides opportunities for rational structure-based drug design approaches to improve existing or obtain novel antibiotics helpful in combating bacterial resistance[4]. However, structural determination by X-ray crystallography only provides static views of the binding processes but does not reveal the dynamics involved with antibiotics binding, or the energetic determinants of binding. Theoretical and computational approaches such as molecular dynamics (MD) simulations in combination with free energy calculations[5] are suitable to fill this gap.

In the present study, we aim at investigating the determinants of binding of the oxazolidinone class; the only synthetic antibiotic class to have entered the market during the last 40 years[6]. Structures of linezolid, the first approved oxazolidinone antibiotic, in complex with the large ribosomal subunits from *H. marismortui* (H50S)[2] and *D. radiodurans* (D50S)[3] have been solved recently. Here, we investigate linezolid, its derivative radezolid, and the structurally related oral anticoagulant drug rivaroxaban in complex with H50S by means of MD simulations to obtain insights into the determinants of binding of this antibiotic class (Fig. 1).
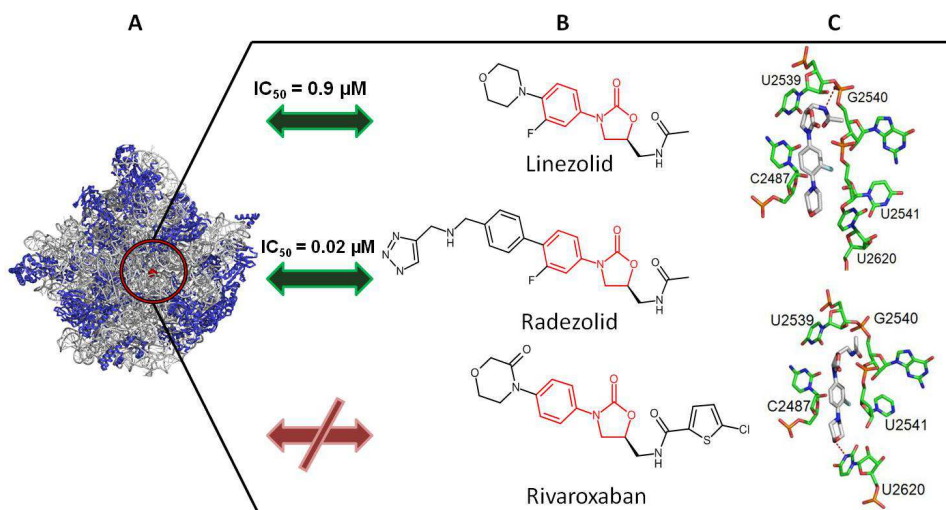
Figure 1. The binding site of oxazolidinones inside the large ribosomal subunit. (A) Crown view of H50S; proteins are shown in blue and RNAs in light grey. (B) The chemical structures of linezolid (top), radezolid (middle), and rivaroxaban (bottom); the phenyl-oxazolidinone core is colored in red. (C) Enlarged view of the linezolid binding site with key residues and respective hydrogen bonds formed (dotted lines) in H50S (top) and D50S (bottom).

2 Methods

The starting structures used for the MD simulations were derived from the 2.7 Å resolution X-ray structure of linezolid bound to H50S (PDB code 3CPW)[2] by structurally modifying linezolid to radezolid and rivaroxaban, respectively. All MD simulations (around 800,000 atoms each) were performed using the AMBER 10 suite of programs with the ff99SB force field and the TIP3P water model. All three simulations reached a length of 50 ns of which snapshots saved at 20 ps intervals were used for analysis.

3 Results

As a first step we investigated the local interactions of the oxazolidinone derivatives formed within the binding site. This analysis showed that the hydrogen bond between the acetamide NH of radezolid and the phosphate group of G2540 (H50S numbering used throughout) persists during the whole simulation, whereas it breaks in the case of linezolid after 4 ns and only forms for a few nanoseconds in the case of rivaroxaban (between 28 and 32 ns) (Fig. 2, (1)).

In a second step, the aromatic stacking interactions between the nucleobase of U2539 and the oxazolidinone core (Fig. 2, (2)) as well as between C2487 and the phenyl rings of the oxazolidinone derivatives (Fig. 2, (3)) were investigated. The aromatic interactions persist during the MD trajectories in both the radezolid and linezolid complexes but break after 30 ns in the case of the rivaroxaban complex. This finding is in agreement with

174
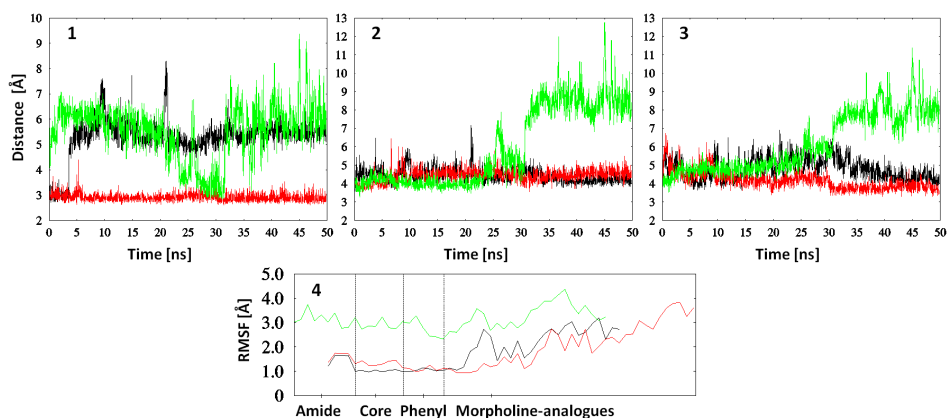
Figure 2. Time-course of non-covalent interactions: (1) Hydrogen bond formation between the ligands' acetamide NH and the phosphate group of G2540 and of aromatic interactions between (2) the oxazolidinone core and the nucleobase of U2539 as well as (3) between the phenyl ring and the nucleobase of C2487, respectively. (4) Atomic fluctuations (RMSF) of the atoms of oxazolidinones during the MD trajectories of H50S complex structures with linezolid (black), radezolid (red), and rivaroxaban (green).

the observed ligand movement inside the binding site (Fig. 2, (4)): While the core of rivaroxaban shows large fluctuations (RMSF $\sim$ 3.0 Å), the core of linezolid and radezolid are largely immobile (RMSF $<$ 1.5 Å).

Finally, the fluctuations of the binding site residues during the MD trajectories were investigated. Again, a similar behavior was observed in the case of the linezolid and radezolid simulations, where both binding sites undergo moderate movements (RMSF $\sim$1 Å). In contrast, the binding site of the rivaroxaban complex shows much more pronounced fluctuations (Fig. 3) that reflect the movement observed for the ligand. In that case, the largest movement is observed for the nucleobase of U2620.
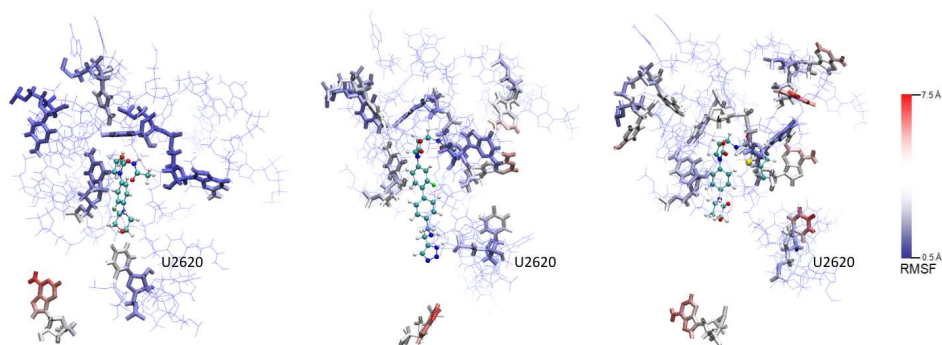
Figure 3. Color-coded representation of atomic fluctuations obtained from the MD trajectories for the binding site residues of the H50S complex structures with linezolid (left), radezolid (middle), and rivaroxaban (right). Red color indicates high and blue color low RMSF values.

175

4 Discussion

The present study underlines the importance of computational studies to analyze antibiotics binding to the ribosome. While certain aspects of binding of linezolid to H50S can be understood based on the available complex crystal structures, information about the binding determinants of radezolid and rivaroxaban to H50S has remained elusive so far.

To investigate the determinants of binding of antibiotics at the ribosomal structures, we have performed all-atom MD simulations of large ribosomal subunits in explicit solvent. The analysis of hydrogen bonds and aromatic interactions implicates a stronger binding of radezolid over linezolid, which is in agreement with the experimental data. Surprisingly, while the chlorothiophene moiety of rivaroxaban can be well accommodated in the H50S crystal structure, the MD simulations reveal a pronounced movement of this ligand, which is accompanied by a loss of the initial hydrogen bonding and aromatic interactions. This may be the reasons why no inhibition of protein synthesis was detected so far for rivaroxaban.

In contrast, the binding of linezolid and radezolid is mainly stabilized by the acetamide moiety as it is involved in making hydrogen bonding interactions with the RNA. In agreement, all recent linezolid derivatives preserve this moiety, which shows that any change in the group can be detrimental to the affinity. However, the morpholino group does not form apparent interactions and can be substituted with other groups without leading to a loss in inhibitory activity. This is supported by SAR studies on linezolid and radezolid by others. The triazole moiety for both linezolid and radezolid forms weak hydrogen bond interactions with the nucleobase of U2620. This interaction stabilizes the conformation of U2620, thus arresting it in a nonproductive conformation, and thereby prevents the correct positioning of the P-tRNA. In the case of rivaroxaban, U2620 undergoes large movements during the MD simulation and forms no stable interaction.

By means of MD simulations, we were able to identify key structural features required for the strong binding of radezolid to H50S. Also, we were able to identify structural reasons why rivaroxaban does not act as an antibiotic. Currently, we are in the process of investigating the contributions of the binding site residues to the relative binding free energies of the three oxazolidninone derivatives. Together with the structural information gained so far, this information will be used to further understand antibiotics selectivity and resistance. Furthermore, we expect our findings to provide a basis for designing improved antibiotics.

Acknowledgments

References

1. Yonath, A., *Antibiotics targeting ribosomes: resistance, selectivity, synergism, and cellular regulation*, Annu. Rev. Biochem., Vol. 74, 649-679, 2005.
2. Ippolito, J.A. et. al., *Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit*, J. Med. Chem., Vol. 51, 3353-3356, 2008.
3. Wilson, D.N. et.al., *The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning*, Proc. Natl. Acad. Sci., Vol. 105, 13339-13344, 2008.
4. Fulle, S., Gohlke, H., *Statics of the ribosomal exit tunnel: implications for co-translational peptide folding, elongation regulation, and antibiotics binding*, J. Mol. Biol., Vol. 387, 502-517, 2009.
5. Gohlke, H., Case, D.A., *Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf*, J. Comp. Chem., Vol. 25, 238-250, 2004.
6. Moellering, R.C., *Linezolid: the first oxazolidinone antimicrobial*, Ann. Intern. Med., Vol. 139, 135-142, 2003.

Structural Influence on the Binding Behavior in Polymorphisms and Resistant Mutants of the NS3/4A Serine Protease of the Hepatitis C Virus

Sabine Schweizer[1], Christoph Welsch[2], and Iris Antes[1]

[1] Department of Life Sciences and Center for Integrated Protein Science Munich
Technical University Munich, 85354 Freising-Weihenstephan, Germany
E-mail: {sabine.schweizer, antes}@wzw.tum.de

[2] Department of Medicine, Lineberger Comprehensive Cancer Center
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA
E-mail: christophwelsch@gmx.net

Hepatitis C virus (HCV) infections affect 3% of the world's population and are a serious global health problem. Rapid drug resistance development of HCV against direct-acting antiviral (DAA) agents is regarded as a major problem in medical treatment. For rational drug design, it is therefore an important challenge to gain a detailed understanding of the molecular mechanisms conferring resistance. We present computational studies of natural and drug-resistance related viral variants of the HCV NS3/4A serine protease which is considered a most promising drug target in DAA development. The structural basis of the binding properties of the wild type protease and the resistance behavior of the protease inhibitor-resistance related viral variants V55A/I and R155K/Q/T are analyzed on the basis of molecular dynamics simulations. Our studies indicate two main effects leading to a reduced susceptibility of HCV protease inhibiting compounds and provide an explanation of the experimental resistance data on a molecular basis.

1 Introduction

According to estimates of the world health organization, worldwide currently about 170 million people are infected with the hepatitis C virus (HCV)[1]. Most patients develop chronic infection and bear an increased risk of suffering severe life-threatening long-term afflictions[1]. HCV drug and vaccine development is a most challenging task because the virus circulates as an ensemble of genetic variants, so-called quasispecies that represent a main obstacle for drug design[1]. The hepatitis C NS3/4A serine protease is regarded as a highly promising target for direct-acting antiviral (DAA) agents. A new generation of direct-acting linear ketoamide inhibitors shows encouraging results in recent clinical trials[2], but owing to the quasispecies population the virus is able to rapidly develop drug resistance under selective pressure, which is regarded as major problem in future medicinal treatment[3,4]. Thus, it is especially with regard to rational drug design of great interest to gain a detailed understanding of the resistance mechanisms. Typically, rather small structural changes are involved in resistance development. This is in particular true for viral escape mechanisms as the virus must conserve its functionality even in resistant mutants and cannot undergo large conformational changes.

In the following, we present MD studies of the structural influence of V55 and R155 point mutations on the binding properties and resistance behavior of the HCV NS3/4A serine protease and the natural inhibitor-resistance related V55A/I and R155K/Q/T variants.

2 Methods

The crystal structures with the PDB IDs 2OC8[5] (wild type (WT) with covalently bound inhibitor boceprevir) and 1A1R[6] (apo WT structure) served as structural basis. The starting structures for the variants were obtained by mutating the side chains V55 and R155, respectively, using the software tool IRECS[7]. Energy minimizations and MD simulations were performed using the GROMACS[8] program package together with the GROMOS96 53a6 force field parameter set[9]. Topology parameters for the ligand were obtained from the PRODRG server[10]. To specify the covalent linkage between S139 and the inhibitor, a deprotonated serine was defined based on the parameters of the corresponding standard residue of the employed force field and suitable parameters for bond length and angles were added in the topology file to define the covalent bond. Each system was neutralized and embedded in a periodic cubic water box. The water molecules were described by the simple point charge (SPC) explicit water model. For the evaluation of non-bonded interactions a cutoff of 1.4 nm was used and electrostatic interactions were calculated with the particle-mesh Ewald (PME) method. Bond lengths were constrained with the LINCS algorithm. The Berendsen weak coupling method was applied to keep temperature and pressure constant with a temperature coupling relaxation time of 0.1 ps, a pressure coupling constant of 1.0 ps, and a compressibility of $4.5 \cdot 10^{-5}$. The energy minimized structures were stepwise heated from 0 to 300 K over 460 ps. During the heat-up procedure, a time step of 1 fs was applied together with position restraints up to a temperature of 200 K. For the subsequent equilibration over 20 ns at 300 K, a time step of 2 fs and a constant pressure of 1 atm was used.

3 Results and Discussion

The binding pocket of the HCV NS3/4A serine proteases is characterized by a catalytic triad comprising residues S139, H57, and D81. Linear ketoamide inhibitors, mimicking the natural substrate, bind covalently to the catalytically active S139, while the first generation of HCV protease inhibitors, macrocyclic compounds, binds through non-bonded interactions. The mutated residue V55 is not directly involved in inhibitor binding, but is situated adjacent to S139 screened from the binding site by the catalytically active H57 and D81. In contrast, residue R155 is exposed to the binding site surface and forms a salt bridge network with its neighboring residues D81, D168, and R123. Thus, different types of resistance mechanisms can be expected upon mutation of V55 and R155.

In order to analyze the structural basis of the mutants' influence on the binding behavior, we performed MD simulations for the WT and the V55A/I variants with and without the inhibitor boceprevir to study the influence of the mutation on the binding site geometry in the bound and empty state. Furthermore, simulations were carried out for R155K/Q/T protease variants without inhibitor. For comparison, we additionally performed calculations for the apo WT protease to rule out possible artifacts by deleting the inhibitor from the bound complex structure.

3.1 V55 Variants

In comparison to the WT, the inhibitor-resistant variant V55A shows a constriction of the binding pocket around the binding site S139 (see Fig. 1 (a)).
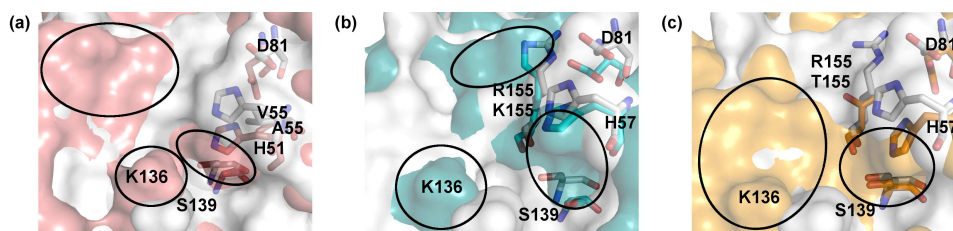
Figure 1. Superposition of the binding sites of the apo WT structure (grey) and the unbound variants (a) V55A (pink), (b) R155K (cyan), and (c) R155T (orange). Protruding surface areas of the variants into the binding cavity are highlighted.

This is found for both the bound and the unbound state and is reflected in C_α distances between S139 and the nearby residues within the binding site region. Thus, the narrowed binding pocket reduces the susceptibility as it hampers the inhibitor molecule to bind covalently to the protease. In contrast, similar conformational restrictions are not observed for the V55I variant which correlates well with the experimental data.

3.2 R155 Variants

The analysis of the MD simulations of the R155K/Q/T variants suggests two different structure-based effects that reduce the susceptibility: First, all mutations affect the stability of the electrostatic network within the binding site. The number of salt bridges is reduced in the variants (R155K: # = 7, R155Q: # = 3, R155T: # = 3)[a] compared to the WT (apo WT: # = 8). This decrease weakens interactions with macrocyclic inhibitors. Second, mutations to lysine and threonine, respectively, cause conformational changes within the binding pocket similar to the effects observed for the V55A variant (see Fig. 1 (b) and (c)). These sterical constrictions hamper ketoamide inhibitors to access the binding site.

4 Conclusion

Our studies provide a deeper insight into the molecular resistance mechanisms of the HCV NS3/4A serine protease and an explanation of experimentally observed binding affinities of the natural inhibitor-resistance related protease variants V55A/I and R155K/Q/T.

Acknowledgments

---

[a]The number of salt bridges (denoted as #) was obtained by analyzing a 10 Å sphere around residue 155 within the central cluster structure using VMD with a oxygen-nitrogen cut-off of 10 Å.

References

1. E. Barnes, *Viral Cancers - Hepatitis C Virus*, Tech. Rep., World Health Organization, 2010, and references therein.
2. S. Margeridon-Thermet and R. W. Shafer, *Comparison of the Mechanisms of Drug Resistance among HIV, Hepatitis B, and Hepatitis C.*, Viruses, **2**, 2696–2739, 2010.
3. K. P. Romano, A. Ali, W. E. Royer, and C. A. Schiffer, *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding.*, PNAS, **107**, 20986–20991, 2010.
4. T. Shimakami, C. Welsch, D. Yamane, D. R. McGivern, M. Yi, S. Zeuzem, and S. M Lemon, *Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus.*, Gastroenterology, **140**, 667–675, 2011.
5. A. J. Prongay et al., *Discovery of the HCV NS3/4A Protease Inhibitor (Sch 503034) II. Key Steps in Structure-Based Optimization*, J. Med. Chem., **50**, 2310–2318, 2007.
6. J. L. Kim et al., *Crystal Structure of the Hepatitis C Virus NS3 Protease Domain Complexed with a Synthetic NS4A Cofactor Peptide*, Cell, **87**, 343–355, 1996.
7. 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 Sci., **16**, 1294–1307, 2007.
8. B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput., **4**, 1549–9618, 2008.
9. C. Oostenbrink, A. Villa, A. E. Mark, and 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.
10. A. W. Schuettelkopf and D. M. F. van Aalte, *PRODRG - a tool for high-throughput crystallography of protein-ligand complexes*, Acta Cryst. D, **60**, 1355–1363, 2004.

Improving Internal Peptide Dynamics in the Coarse-grained MARTINI Model: Application to Amyloid and Elastin Peptides

Mikyung Seo[1], **Sarah Rauscher**[2], **Régis Pomès**[2], **and D. Peter Tieleman**[1]

[1] Department of Biological Sciences and Institute for Biocomplexity and Informatics
University of Calgary, Calgary, Alberta, Canada
E-mail: {miseo, tieleman}@ucalgary.ca

[2] Department of Biochemistry, University of Toronto and
Molecular Structure and Function, Hospital for Sick Children, Toronto, Ontario, Canada
E-mail: sarah.rauscher@utoronto.ca, pomes@sickkids.ca

The recently developed MARTINI coarse-grained model reproduces a wide range of lipid properties as well as lipid-protein interactions for rigid proteins. Protein folding and aggregation however often involve significant transitions between secondary structures and hence require that the proteins be flexible during the simulations. We present recent advances in our extension of the MARTINI model to more accurately describe the internal flexibility of peptides by introducing in the energy function a term that accounts for the dihedral potentials on the peptide backbone. We assess the performance of the resulting model by calculating structural properties and comparing them with the outcome of atomistic simulations. The improved model is also applied to investigate the self-assembly of β-sheet forming peptides. We characterize the conformation of peptides and follow their self-aggregation in water and at the water-octane interface.

1 Introduction

Many phenomena in biological systems such as protein folding and peptide aggregation occur on long time scales. Coarse-grained (CG) models present an attractive alternative to the traditional atomistic simulations since they offer the possibility of investigating complex biological processes over larger length and longer time scales at a reduced level of detail. In CG models, groups of atoms are represented as beads, thus lowering the total number of particles in the system. The reduced number of degrees of freedom and the use of smoother interaction potentials make the model computationally very efficient. Marrink and co-workers developed one such CG model for simulation of lipids and surfactants, coined the MARTINI force field[1,2]. The MARTINI force field was later extended to proteins[3] and it has been successfully applied in numerous studies in the past few years. However, there are certain limitations in the current implementation of the MARTINI force field. Changes in the secondary structure of protein are not modeled, as backbone bonded parameters are currently dependent on pre-defined secondary structures. Therefore, biological processes in which the folding and unfolding of secondary structures play an important role cannot be easily studied with the current MARTINI force field.

We report here the extension of the CG MARTINI model to improve the treatment of the backbone flexibility of peptides achieved by reparameterizing dihedral angle potentials and its application to amyloid and elastin-like peptides in order to model the main events during self-assembly.

2 Systems and Methods

We used two sets of amyloid and elastin-like peptides: SNNFGAIL, $(GA)_4$, $(GV)_4$, GVGVAGVG, GVGVGGVG, GVAGVAGV, GVGVAGGV, GVGGVGGV (amyloid), GVGVPGVG, GVPGVPGV, GVAPGVGV, GVGGVPGV (elastin).

2.1 Parameterization of the Dihedral Angle Potential

To develop dihedral angle potentials of the peptide backbone in CG representation, extensive conformational sampling was performed by atomistic simulations[4] and various structural properties calculated from these atomistic trajectories were used as reference for our CG simulations. The bond and angle interactions were modeled by the harmonic energy functions in the original MARTINI peptide force field[3]. The dihedral angle was defined by four consecutive backbone beads, and the dihedral angle distributions were calculated for every quartet of residues possible for each peptide. Using the center of mass of the group of atoms corresponding to CG beads, the distributions of the dihedral angles were calculated for every peptide from atomistic trajectories. Then, the corresponding potentials of each dihedral angle were extracted from the normalized probability distributions by using the Boltzmann inversion procedure. In the final step, potential energies were fitted to a sum of cosine and sine terms:

$$V_d = \sum_{i=1}^{4}[C_i \cos(i\phi) + S_i \sin(i\phi)] \tag{1}$$

where ϕ is the dihedral angle and C_i and S_i values are the force constants.

3 Results and Discussion

The performance of our model was assessed by comparing the backbone beads (BB) contact maps with their counterparts from atomistic simulations. In Fig. 1, we show the BB contact maps of two selected peptides, GVAPGVGV and $(GV)_4$, obtained from CG and atomistic simulations. The contact maps show the probability that a contact is present, as a percentage of the total simulation time. For all cases, the new model reproduced the most populated contact of atomistic results: the A3-V6 contact in peptide GVAPGVGV and the G3-V6 contact in peptide $(GV)_4$. Several other contacts with lower populations were also reproduced with the new model. Our results indicate that the performance of the model was significantly improved with the addition of new dihedral potentials.

Next, we examined whether the improved description of the backbone dihedrals is adequate for studying aggregation processes without additional parameter adjustments. CG simulations of 64 SNNFGAIL monomers were performed both in water and in the presence of a hydrophobic octane phase. It has been reported that monomers of simple β-sheet forming peptides in water and at the interface adopted many different conformations and interconverted over the course of the simulations[5]. To analyze the conformation of the peptides within the aggregates we calculated the distribution of end-to-end distances, d_{ee} for aqueous and adsorbed peptide monomers at the interface (Fig. 2). Three conformation types were defined based on d_{ee}: short (S) for $d_{ee} < 0.65$ nm, intermediate (I) for
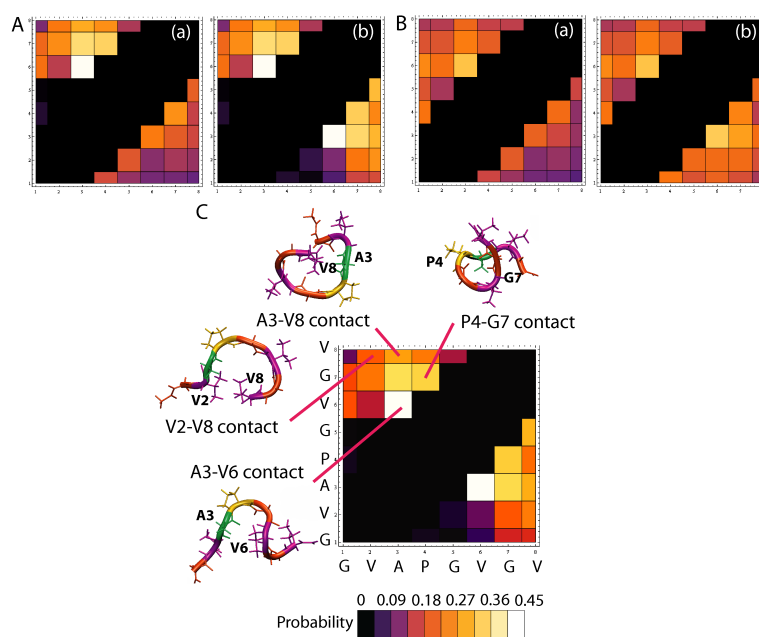
Figure 1. Backbone beads contact maps at 305 K. Each square in the matrix (i,j) corresponds to a contact between the backbone beads of residue i and j.; GVAPGVGV (A); $(GV)_4$ (B); Representative snapshots of GVAPGVGV showing the presence of significantly populated contacts. In the color scheme, each color represents a range of probabilities of contact formation (C). On each map, the atomistic map lies above the diagonal, and the corresponding CG map, obtained from the simulations using either original MARTINI model (a) or the new model with the addition of dihedral potentials (b), lies below the diagonal.
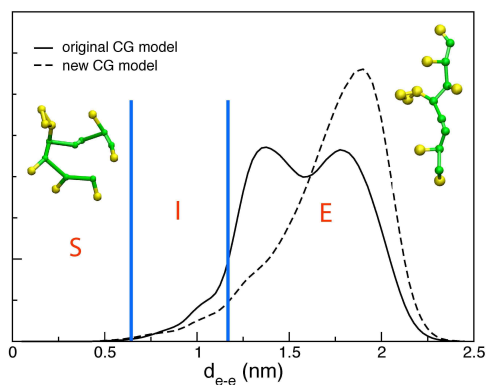
Figure 2. Distribution of end-to-end distance of 64 SNNFGAIL at water-octane interface using the original and new CG models. Vertical lines at $d_{ee} = 0.65$ and 1.2 nm highlight the boundaries separating short, intermediate and extended conformations. Representative snapshots of the peptide in short and extended conformations are shown.

$0.65 < d_{ee} > 1.2$ nm and extended (E) for $d_{ee} > 1.2$ nm. End-to-end distributions calculated from atomistic trajectories[5] showed that peptides at the interface preferred to adopt an extended conformation compared to aqueous peptides, indicating that the adsorption of peptides at the interface displaced the conformational equilibrium of the peptides in favor of stretched conformers. Comparison of the distributions of d_{ee} obtained from the original MARTINI model and the new model reveals that both CG models favor extended structures. With the new model, however, the distribution of peptides is shifted even further towards extended conformations resembling the β-strands at the water-octane interface during peptide aggregation. Based on these observations and those made from analysis of atomistic trajectories[5], we conclude that the inclusion of torsional backbone flexibility in the new model provides a superior description of peptide conformational preferences during aggregation at the interface.

4 Concluding Remarks

The present work introduced a new approach for modeling the backbone flexibility of proteins as extension of the CG MARTINI model. We applied this approach to amyloid and elastin-like peptides. Peptide contact maps showed a significant improvement when torsional flexibility was allowed in the new model. In addition, our new CG model was able to reproduce the displacement in the conformational equilibrium in favor of extended structures during aggregation at the interface without further parameter adjustment, making the improved MARTINI model suitable for biological processes involving peptide flexibility over long time scales.

Acknowledgments

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. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*, J. Phys. Chem. B **111**, 7812–7824, 2007.
3. L. Monticelli, S. K. Senthil, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, *The MARTINI Coarse-Grained Force Field: Extension to Proteins*, J. Chem. Theory. Comput. **4**, 819–834, 2008.
4. S. Rauscher, C. Neale, and R. Pomès, *Simulated Tempering Distributed Replica Sampling, Virtual Replica Exchange, and Other Generalized-Ensemble Methods for Conformational Sampling*, Proteins **5**, 2640–2662, 2009.
5. A. Nikolic, S. Baud, S. Rauscher, and R. Pomès, *Molecular Mechanism of β-sheet Self-Organization at Water-Hydrophobic Interfaces*, Proteins **79**, 1–22, 2011.

Predicting Cartilage Inflammation Due to NF-kB Activation Using a Multiscale Finite Element Model of the Knee

Vickie B. Shim[1], Peter J. Hunter[1], Peter Pivonka[2], and Justin W. Fernandez[1]

[1] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand
E-mail: v.shim@auckland.ac.nz

[2] University of Western Australia, Perth, Australia

The initiation of osteoarthritis has been linked to the onset and progression of pathologic mechanisms at the cartilage-bone interface, especially the pathway of NF-kB activation, which up-regulates a number of pro and anti inflammatory genes. Moreover, this degenerative disease involves cross-talk between the cartilage and subchondral bone environments so an informative model should contain the complete complex. We have developed a multiscale model using the open-source ontologies with cartilage and bone descriptions at the cellular, micro and macro levels to effectively model the influence of whole body loadings at the macro level on tissue and cell level processes that determine bone and cartilage remodelling. We evaluate the framework by linking a common knee injury (ACL deficiency) to proinflammatory mediators as a possible pathway to initiate osteoarthritis. This framework provides a virtual bone-cartilage tool for evaluating hypotheses, treatment effects and disease onset to inform and strengthen clinical studies.

1   Introduction

Osteoarthritis (OA) is a debilitating disease that causes widespread physical morbidity and impaired quality of life . Traditionally, OA has been regarded as the wear and tear of cartilage. However, recent evidence suggests that; (i) OA is not just a disrupted cartilage mechanism but a breakdown of the whole cartilage-bone complex, hence, bone and cartilage should be treated together due to their intrinsic inter-relationship; and (ii) OA exists in the highly metabolic and inflammatory environments of adipose tissue, which control joint degradation or local inflammatory processes by secreting various proinflammatory cytokines[1]. Traditional biomechanical studies have focused on macro level factors mainly, such as cartilage thinning or material property change. The question of whether degeneration in cartilage is initiated by changes in articular cartilage or the underlying subchondral bone is difficult to answer when one considers them in isolation. Therefore, our modelling approach aims to have a holistic view of bone-cartilage physiology with information viewed collectively rather than in a fragmented approach. This study employs a multiscale framework developed as part of the Physiome Project[2,3]. In order to integrate these modelling scales an efficient set of numerical tools is necessary to link, store models and employ different methodologies at different scales. Here we illustrate this structure for an important musculoskeletal problem, OA induced at the cartilage-bone interface.

The aim of this study is to present a comprehensive multiscale finite element (FE) framework of the cartilage bone junction in order to analyze the initiation of OA. This is accomplished by numerically linking information across spatial scales from a whole

continuum knee model down to a model of cell mechanics described using ordinary differential equations. To initiate this process we adopt a well-known condition, anterior cruciate ligament deficiency (ACLD) that leads to abnormal mechanical loading at the knee as a possible precursor[4]. We hypothesize that when abnormal loading is applied, such as with ACLD gait, bone remodelling will modify micro architecture and ultimately influence peak cartilage strains. These strains will then induce proinflammatory cytokines to degrade cartilage quality at the whole organ level.

2 Methods

An anatomically-based knee model developed to investigate joint loading in gait was the starting point[5]. This model consisted of the femur, tibia, ligaments, menisci and cartilage and the joint forces and torques were estimated using inverse dynamics. The geometries were developed using high-order cubic Hermite elements, fitted to the Visible Human dataset and customized to a subject using free-form deformation methods[6]. The model assembly is detailed in AnatML[3], which describes the anatomical linkages for the whole human body and acts as a store for the anatomical geometries. To model the loads experienced by the knee we simulated gait, the most common human task. Two key loading points of the gait cycle were identified, heel-strike and contralateral toe-off to load the cartilage. Within the cartilage we embedded micro models of the cartilage bone interface representing 1 mm cubes to capture the micro deformations. These were placed at key locations across the cartilage surface to capture the spatial variation of deformation. The linking between the macro and micro models was integrated within our finite element software CMISS (`www.cmiss.org`). This package is open-source and facilitates the numerical linkages between the modelling scales. To simulate a condition known to induce cartilage degeneration we then simulated an anterior cruciate ligament deficiency (ACLD). The ACL was removed predicting increased anterior movement of the tibia. At the micro-scale a detailed micro FE model was developed, which comprised of cartilage, a zone of calcified cartilage (ZCC) and subchondral bone, all treated as deformable adapted from the imaging study by Zizak et al.[7]. During each gait cycle the strain boundary conditions were passed to an embedded micro model of the cartilage bone interface. This micro model computed the local strain field, which was homogenized and passed back to the macro cartilage model in a two way feedback loop. The micro model was informed by two cell models. A bone remodelling algorithm[8] based on the RANK-RANKL-OPG pathway, which predicts the number of active osteoblasts (to deposit bone) and osteoclasts (to absorb bone) coded in CellML[9]. A cartilage damage prediction model was developed based on the work of Nam et al.[10], which gives a quantitative description of the action of nuclear factor-kappa B (NF-B) signaling cascade under mechanical stimulation. Peak cartilage strains were used as excitation of the protein complex IB kinase (IKK) which activates the NF-B pathway, leading to induction of a number of proinflammatory genes.

3 Results

The previously validated whole knee model[5] was consistent with cartilage deformations previously reported [24]. Fig. 1 illustrates a slice of the 3D micro FE model and an illus-
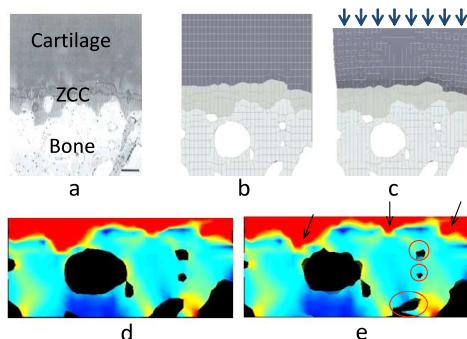
Figure 1. (a) Electron microscope image showing cartilage, zone of calcified cartilage and subchondral bone; (b) 2D slice of undeformed FE mesh and (c) compressed FE mesh with 100m loading; (d) healthy von Mises strain and bone architecture; (d) ACLD von Mises strain and remodelled architecture after 3 months of bone evolution.
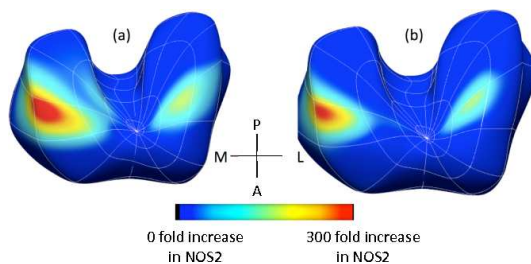

Figure 2. Inflammatory response measured in terms of the expression level of nitric oxide synthase (NOS2) for (a) ACLD and (b) normal loadings.

trative micro cartilage deformation of 100 m. It is clear that the subchondral bone architecture creates regions of strain shielding leading to a spatially varying strain excitation signal. This produced an increase in the peak strains observed in the cartilage, which when coupled to the cartilage cell model increased the production of inflammatory cytokines as shown in Fig. 2, which shows the pattern of proinflammatory initiation across the cartilage in the remodelled knee for the ACLD loading (Fig. 2a) and normal loading (Fig. 2b). In the anterior cartilage, primarily in the medial compartment, the model predicted an inflammatory response highlighting sites of likely degeneration. What is most interesting from our model prediction is that even after a return to normal loading (Fig. 2b) inflammation continued. This indicates that once the pathologic state is initiated, the degenerative cycle may not be reversible.

4 Discussion

We have presented a numerical framework showing the linkages between AnatML, FieldML and CellML. This has been used to evaluate a possible mechanism for the ini-

189

tiation of OA by considering the coupling of bone and cartilage. This was established by considering the excitation for bone remodelling being influenced by the complex subchondral bone architecture to induce a spatially varying remodelling pattern in the cartilage bone tidemark. This, in turn, leads to changes in peak cartilage strains that can induce proinflammatory responses. This is consistent with previously reported work that showed altered mechanical loads can lead to loading in more fibrillated and thinner regions of cartilage, which may initiate cartilage breakdown[4]. The presented modelling framework is a web-based open-source repository of models, which contributing scientists can easily use and disseminate their findings. The findings in this study may explain why altered or traumatic subchondral bone loadings can initiate a degenerative cycle in the cartilage matrix. In conclusion, we have demonstrated the first attempt at linking three scales together to explain a possible mechanism for the initiation of OA. The use of the Physiome Project infrastructure has made linking the scales more efficient, and the markup languages (AnatML, FieldML and CellML) provide a structure for easy model storage and dissemination.

References

1. R. M. Aspden, B. A. Scheven, and J. D. Hutchison, *Osteoarthritis as a systemic disorder including stromal cell differentiation and lipid metabolism*, Lancet **357**, 1118-20, 2001.
2. P. J. Hunter, and T. K. Borg, *Integration from proteins to organs: the Physiome Project*, Nat Rev Mol Cell Biol **4**, 237-43, 2003.
3. P. J. Hunter, *The IUPS Physiome Project: a framework for computational physiologys*, Prog Biophys Mol Biol **85**, 551–69, 2004.
4. T. P. Andriacchi, S. Koo, and S. F. Scanlan, *Gait Mechanics Influence Healthy Cartilage Morphology and Osteoarthritis of the Knee* Journal of Bone and Joint Surgery-American Volume **91A**, 95-101, 2009.
5. V. B. Shim, K. Mithraratne, I. A. Anderson, and P. J. Hunter, *Simulating In-Vivo Knee Kinetics and Kinematics of Tibio-Femoral Articulation with a Subject-Specific Finite Element Model* Proceedings of the IFMBE **25/4**, 2315-2318, 2010.
6. J. W. Fernandez, P. Mithraratne, S. F. Thrupp, M. H. Tawhai, and P. J. Hunter, *Anatomically based geometric modelling of the musculo-skeletal system and other organs* Biomech Model Mechanobiol **2**, 139-55, 2004.
7. I. Zizak, P. Roschger, O. Paris, B. M. Misof, A. Berzlanovich, S. Bernstorff, *Characteristics of mineral particles in the human bone/cartilage interface* J Struct Biol **141**, 208-17, 2003.
8. P. Pivonka, J. Zimak, D. W. Smith, B. S. Gardiner, C. R. Dunstan, N. A. Sims *Model structure and control of bone remodeling: a theoretical study* Bone **43**, 249-63, 2008.
9. D. Nickerson and P. Hunter, *Using CellML in computational models of multiscale physiology* Conf Proc IEEE Eng Med Biol Soc **6**, 6096-9, 2005.
10. J. Nam, B. D. Aguda, B. Rath, and S. Agarwal, *Biomechanical thresholds regulate inflammation through the NF-kappaB pathway: experiments and modeling* PLoS One **4**, e5262, 2009.

Determination of the Potentials of Mean Force for Stretching of $C^\alpha \cdots C^\alpha$ Virtual Bonds in Polypeptides from the *ab initio* Energy Surfaces of Terminally-blocked N-methylacetamide and N-pyrrolidylacetamide

Adam K. Sieradzan[1], Harold A. Scheraga[2], and Adam Liwo[1]

[1] Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland
E-mail: {adasko, adam}@chem.univ.gda.pl

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

Potentials of mean force for deformation of virtual $C^\alpha \cdots C^\alpha$-bonds were determined as functions of virtual-bond length to extend the application of our coarse-grain UNRES force field to handle the cis-trans transition of peptide groups. The potentials were calculated through Boltzmann summation over the energy surfaces of N-methylacetmide and N-acetylpyrrolidyne, which served as model compounds. The respective energy surfaces were calculated on the $C^\alpha \cdots C^\alpha$ virtual-bond length (d) and and the improper dihedral angle H-N-C$\cdots$C$'$ (α) grid at the Møller-Plesset (MP2) *ab initio* level of theory with the 6-31G(d,p) basis set. The energy was minimized with respect to the remaining degrees of freedom. The free-energy differences between the cis and the trans form and the free-energy barriers to cis-trans transition calculated from the PMFs conform well with the experimental data. Test simulations were carried out which denotes a $\sim 10^7$- times speed up due to the reduction of the number of the degrees of freedom.

1 Introduction

All-atom *ab initio* MD simulations with explicit treatment of water are still restricted to small proteins and, therefore, simulations of larger proteins require a corse-grained approach. Such an approach is used in the UNRES (UNited RESidue) force field[1], where the representation of each amino-acid residue is reduced to two interacting sites: a peptide group and a side-chain group. The representation of polypeptide chains is simplified and the number of degrees of freedom is reduced, resulting in speeding up simulations by 3 orders of magnitude with respect to all-atom simulations. In contrast to many coarse-grained force fields, UNRES is based upon physical principles and not statistical analysis of protein data bases[1].

Cis-trans isomerization often is the rate-determining step of protein folding. One example is Ribonuclease A[2] in which cis-trans isomerization of three proline residues (Pro^{93}, Pro^{114}, Pro^{117}) determines the folding rate. Moreover, the cis configuration of a peptide group often has major effect on the structure of a protein. Therefore, the aim of this work was to extend the applicability of the UNRES force field to model cis-trans isomerization of peptide groups. This aim was achieved by calculating the potentials of mean force (PMF) of virtual $C^\alpha \cdots C^\alpha$ bond as functions of virtual bond length (d), fitting them to analytical expressions to this potentials, and implementing the resulting analytical function in UNRES.

2 Methods

N-methylacetmide and N-acetylpyrrolidyne were used as models of a regular and proline-type peptide groups, respectively. To determine the respective PMF's, potential-energy surfaces of these systems were calculated by using the MP2/G6-31(d,p) *ab initio* method. The energy was evaluated on the two-dimensional grid, where one coordinate was the $C^\alpha \cdots C^\alpha$ virtual bond length (d) ranging from 2 Å to 5 Å with a step 0.01 Å and the improper H-N-C-C' angle (α), ranging from -90^0 to 90^0 with a 10^0 step. We also tried to use the C^α-N-C(O)-C^α dihedral angle (ω) as a reaction coordinate; however it does not cover small and large $C^\alpha \cdots C^\alpha$ virtual-bond lengths values. The PMFs were calculated by direct Boltzmann summation; they were computed without (Eq. 1) or with (Eq. 2) including the entropy contribution estimated by using a harmonic approximation.

$$F(d,T) = -\beta^{-1} \ln \sum_\alpha \exp[-\beta e^*(d,\alpha)] \tag{1}$$

$$F(d,T) = -\beta^{-1} \ln \sum_\alpha [\det H^*(d,\alpha)]^{-\frac{1}{2}} \times \exp[-\beta e^*(d,\alpha)] \tag{2}$$

where $F(d,T)$ denotes potentials of mean force ; $\beta = 1/RT$ where R is universal gas constant and T is the absolute temperature, $e^*(d,\alpha)$ is the energy at the (d,α) grid point, the superscript "*" denotes value optimized in all degrees of freedom except d and α, $H^*(d,\alpha)$ is the Hessian matrix at the (d,α) grid point.
The PMFs were fitted to an analytical expression given by Eq. 3

$$U(d,T) = -\beta^{-1} \ln \left\{ \exp\left[-\beta \left(\tfrac{1}{2} k_{cis}(d - d^0_{cis})^2 + V_{0cis} + S_{cis}(T - T^0) \right) \right] \right.$$
$$\left. + \exp\left[-\beta \left(\tfrac{1}{2} k_{trans}(d - d^0_{trans})^2 + V_{0trans} + S_{trans}(T - T^0) \right) \right] \right\} \tag{3}$$

where $U(d,T)$ is potential of virtual $C^\alpha \cdots C^\alpha$ bond deformation in UNRES force field k_{cis} (k_{trans}) is force constant for the cis (trans) form, V_{0cis}, (V_{0trans}) is the reference free energy of the cis (trans) peptide bond, S_{cis} (S_{trans}) is first derivative of the free energy of unstrained cis (trans) peptide bond in temperature, $T_0 = 298^0$ K, $d^0_{cis}(d^0_{trans})$ is the equilibrium value of virtual $C^\alpha \cdots C^\alpha$ bond length for the cis (trans) configuration.

This expression behaves as a harmonic potential in the neighbourhood of the cis- or trans-configuration, respectively and includes the dependence of the effective potential on temperature. Marquardt's method[3] was used as the fitting procedure. The PMFs were calculated and fitted on a grid from T = 270^0 K to T = 350^0 K with 5^0 K step size.

After implementation of Eq. 3 in the UNRES force field, four sets of 512 independent Langevin dynamics simulations were run for the Gly-Gly and Gly-Pro dipeptides. The simulation temperatures were set at 300^0 K, 320^0 K, 340^0 K and 360^0 K, respectively. Each trajectory consisted of 10^9 steps with 4.89 fs time-step length. To speed up simulations, the friction of water was set to 10^{-2} normal value. All simulations were started from the trans conformation ($d_{C^\alpha \ldots C^\alpha} = 3.8$ Å). For each series of simulations the fraction of cis conformation averaged over all 512 trajectories was calculated as a function of time and a

first-order kinetic equation was fitted (Eq. 4) to the simulation data.

$$x_{cis} = \frac{k_1}{k_1 + k_{-1}} \{1 - \exp[-(k_1 + k_{-1})t]\} \tag{4}$$

where x_{cis} denotes molar fraction of the cis conformation and k_1 and k_{-1} are forward and backward reaction rate constants. The obtained reaction rate constants were compared with those computed from Kramers' theory of reaction rate (Eq. 5)[4].

$$k = \frac{\Omega}{\eta_{water} \frac{6\pi r}{m_{red}}} \sqrt{2\pi\beta} \times \left\{ \sqrt{m_{red}} \int_{d_1}^{d_2} \exp[\beta U(d)] dq \right\}^{-1} \tag{5}$$

where Ω is frequency of oscillator in the cis or the trans form respectively, η_{water} is the viscosity of water; r is the radius of moving sphere m_{red} is the reduced mass q is the reaction coordinate. $d_1 = d_{trans}$ and $d_2 = d_{cis}$ for the forward reaction and $d_1 = d_{cis}$ and $d_2 = d_{trans}$ for the backward reaction, respectively.

3   Results and Discussion

The calculated PMFs are shown in Fig. 1 for N-methylacetamide and N-acetylpyrrolidyne, respectively. The difference between the PMF values of the cis and trans form is 2 kcal/mol
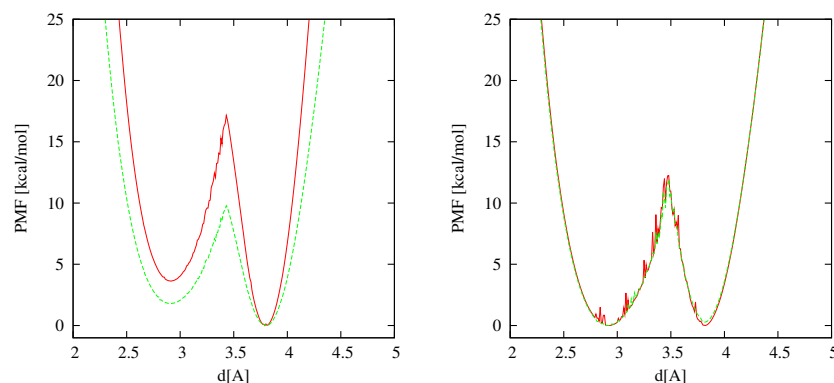
Figure 1. Potentials of mean force without (red) and with (green) entropic contribution for N-methylacetamide (left) and N-acetylpyrrolidine (right) calculated from potential-energy surface by using different schemes at T=298[0]K.

for N-methylacetamide and 0 kcal/mol for N-acetylpyrrolidine, respectively, whereas the free energy barrier to the trans-cis transition equals to 17 kcal/mol and 13 kcal/mol, respectively. After including the entropic contributions calculated from a harmonic approximation, the difference between the PMF at the minima becomes 2 kcal/mol and 0.3 kcal/mol for N-methylacetamide and N-acetylpyrrolidine, respectively, while the free-energy barrier equals 10 kcal/mol and 13 kcal/mol, respectively. Lowering of free energy barrier is caused

by high entropy of the transition state for N-methylacetamide. The value of the free-energy barrier to cis-trans transition obtained for N-methylacetamide is in good agreement with the experimental values of this barrier determined for diglycine, which range from 11 to 13 kcal/mol[5]. Consequently, the PMFs calculated from the (d, α) energy surfaces with including the harmonic contribution to entropy were used to derive the potentials for UNRES (Eq. 3).
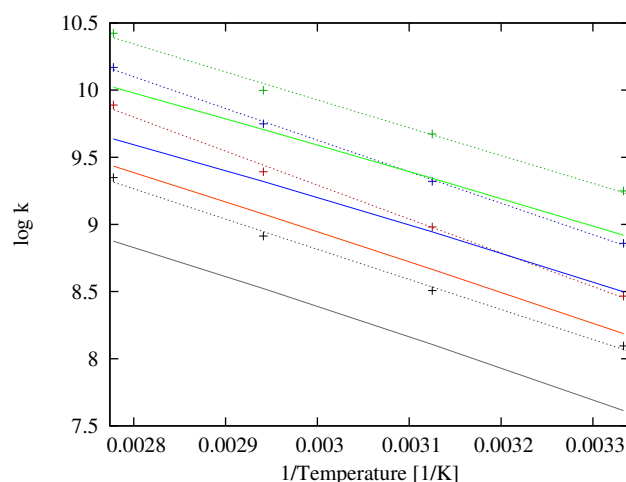

Figure 2. Arrhenius plots of the rate constants of forward and backward reaction of peptide-bond cis-trans isomerization for the Gly-Gly (red: forward reaction, green: backward reaction) and Gly-Pro systems (blue: forward reaction, grey: backward reaction). Crosses: simulation. Dotted lines: Arrhenius fit to values obtained from simulations. Solid lines: values calculated from Kramers' equation.

As shown, the transition from the trans to the cis form is faster for the proline type peptide bond; however, the equilibrium is attained faster for the regular peptide bond (Fig. 2). The trans form dominates for the regular and the cis form dominates for the proline type peptide bond (Fig. 2). The rate constants determined from simulations are in good correlation with those calculated from Kramers' kinetic model (Fig. 2); however they are about 2.5 times greater. The forward (from trans to cis) rate constants are about 10^8 s^{-1} greater at T=300^0 K, compared to experimental values of the order of 10 s^{-1} (Ref. 5). The discrepancy between the UNRES-calculated and experimental rate constants results from the reduction of the number of the degrees of freedom in UNRES. Based on the results of UNRES Langevin dynamics simulations and Kramers' equation analysis, it can be concluded that the reduction of polypeptide chain representation results in a 7 order of magnitude greater rate of cis-trans transition of the peptide bonds compared to experiment, which makes coarse-grained simulations of protein folding, including systems where the cis-trans isomerization is a crucial step, viable.

References

1. A. Liwo, C. Czaplewski, S. Ołdziej, A.V. Rojas, R. Kaźmierkiewicz, M. Makowski, R.K. Muraka and H.A. Scheraga *Simulation of Protein Structure and Dynamics with the Coarse-Grained UNRES Force Field* in Corse-Graining of Condensed Phase and Biomolecular Systems, GA Voth (Ed.), CRC Press, 107–122, 2009.

2. R.W. Dodge and H.A. Scheraga *Folding and Unfolding Kinetics of the Proline-to-Alanine Mutants of Bovine Pancreatic Ribonuclease A* Biochemistry **35**, 1548–1559, 1996.

3. D.W. Marquardt, *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM J. Appl. Math. **11**, 431–441, 1963.

4. H.A. Kramers *Brownian motion in a field of force and the diffusion model of chemical reactions* Physica VII **4**, 284–304, 1940.

5. J.S.W. Holtz, P. Li, and S.A. Asher *UV Resonance Raman Studies of Cis-to-Trans Isomerization of Glycyglycine Derivatives* J. Am. Chem. Soc. **121**, 3762–3766, 1999.

Simulations of Wimley-white Peptides Using the Coarse Grained MARTINI Force Field

Gurpreet Singh and D. Peter Tieleman

Institute for Biocomplexity and Informatics, Dep. of Biological Sciences, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4
E-mail: {gsin, tieleman}@ucalgary.ca

The pentapeptides WLXLL, where X is any of the 20 amino acids, adsorb at the POPC/water interface. Wimley et al. used these peptides to measure the partitioning preferences of amino acids' side chains. Due to their small size and the availability of experimental partitioning free energies, these peptides serve as suitable models for studying peptide adsorption at interfaces using computational methods.

In this study, Wimley-White peptides were simulated at the POPC/water and the cyclohexane/water interfaces using molecular dynamic simulations. The partitioning free energies for amino acids were computed using free energy perturbation methods and are compared with the experimental data. A good agreement was found between the computed and the experimental free energies, with exception of proline and phenylalanine. We also investigate the transferability of dihedral potentials that are derived from the atomistic simulations, to be used in coarse grained simulations.

1 Introduction

Due to rapid progress in computer hardware, our ability to simulate systems containing biomolecules, in atomistic detail, has expanded tremendously. However, sampling all relevant degrees of freedom in a typical system containing hundreds of thousands of atoms remains a challenging task. Biological processes that occur on timescales of milliseconds to seconds are still out of the scope of atomistic simulations.

One approach to increase the reach of simulations towards longer time scales is to simplify the description of the system. Several atoms can be clumped together and represented by a single interaction site, capturing the net effective interactions of the underlying group of atoms, without the loss of significant structural details. The MARTINI model, developed by Marrink and co-workers, represents such an approach[1]. This model was successfully applied to study biological processes such as vesicle fusion and domain formation in vesicles. Originally developed to represent lipids, the MARTINI model was further extended to represent amino acids[2]. The bead mappings for amino acids were chosen to reproduce the cyclohexane/water partitioning free energies of amino acid analogues.

In this study, we compare the partitioning preferences of Wimley-White peptides at the cyclohexane/water interface and the POPC/water interface using the MARTINI force field and highlight key areas where further development in MARTINI model is needed for more accurate representation of lipid/protein interactions.

2 Methods

All the simulations were performed with the GROMACS (version 4.0 and above) software[3]. Peptides of sequence WLXLL, with X as any of the 20 amino acids, were simu-

lated at the cyclohexane/water interface, the POPC/water interface and in bulk water, using MARTINI v2.1[2] for protein and v2.0[1] for lipid parameters. For alchemical simulations, the free energies were computed using both FEP and TI. The MBAR method implemented in the pyMBAR[4] program was used to estimate free energies and uncertainties from FEP data. The detailed simulation setup is described elsewhere[5].

3 Results and Discussions

3.1 Side Chain Distributions

The equilibrium probability densities of the side-chain X at the cyclohexane/water and the POPC/water interface are shown in Fig. 1. The position of the side-chains at the cyclo-hexane/water interface follows their hydrophobicity. The hydrophobic side-chains such as L, V, and I partition deeper towards the cyclohexane, and the polar or the charged side-chains preferentially orient toward the bulk water. The side-chain distributions at both interfaces are quite similar despite the fact that the POPC/water interface is more complex and chemically diverse. On the other hand, partial density profiles of water at the cyclohexane/water and at the POPC/water interface differ significantly. The partial density decreases rapidly from its bulk value in the case of the cyclohexane/water interface, whereas at the POPC/water interface, the decrease is gradual and non-linear, resulting in a broader interface.
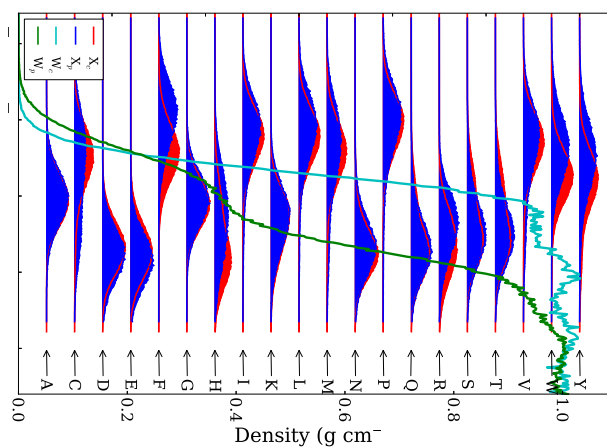
Figure 1. Side-chain COM distributions (X_c and X_p) and partial densities of water (W_c and W_p) at the cyclo-hexane/water and the POPC/water interface. The side-chain distributions are filled. For comparisons, the x-axis was adjusted such that the maximum in the probability density of alanine is at 0 at both the cyclohexane/water and POPC/water interface.

3.2 Partitioning Free Energies

Fig. 2 compares the partitioning free energies calculated using the free energy perturbation method for both the cyclohexane/water and POPC/water interface with that of the exper-
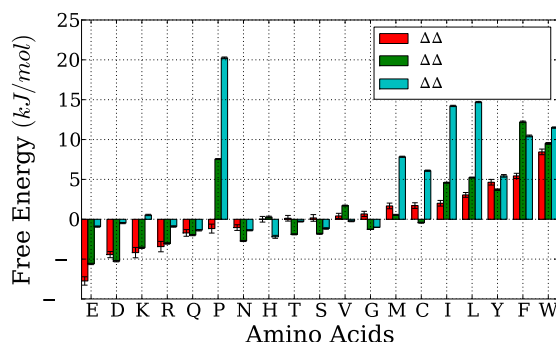
Figure 2. Computed partitioning free energies of side-chains at the POPC/water interface ($\Delta\Delta G_{Calc}(POPC/W)$) and at the cyclohexane/water interface ($\Delta\Delta G_{Calc}(Chex/W)$) as compared to the Wimley-White scale ($\Delta\Delta G_{ww}$).

imental Wimley-White scale. For side-chains with a net charge at pH 7 (D, E, K and R), the free energies were calculated using the polarizable water model of MARTINI. The calculated free energy values at the POPC/water interface are in good agreement with the experimental data, with the exception of P and F. Also, with the exception of V and F, the partitioning free energies are more positive at the cyclohexane/water interface as compared to that of the POPC/water interface.

3.3 Transferability of Dihedral Potential

The current MARTINI model cannot represent the changes in the secondary structure that could occur during the simulations. Even though the protein can be modeled as a helix, coil, turn or bend, once chosen, the secondary structure remains fixed during the course of the simulations. To allow conformational transitions, one approach is to derive the coarse grained dihedral potentials from atomistic simulations. However, such dihedral potentials
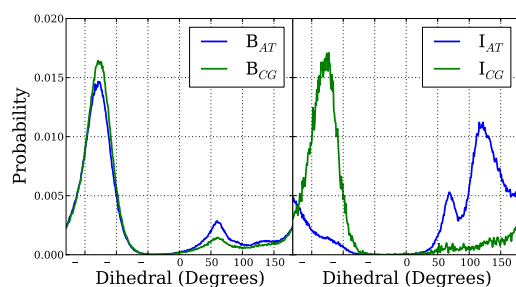
Figure 3. Probability densities of the backbone dihedral angle between last four C_α of the peptide WLELL, in bulk water (Right panel) and at cyclohexane/water interface (Left Panel), obtained from both atomistic (B_{AT} and I_{AT}) and coarse grained simulations (B_{CG} and I_{CG}).

are non-transferable; not only they are specific for the given sequence, they are also specific to the system in which they were obtained. In Fig. 3, B_{AT} and I_{AT} are the probability densities of a dihedral obtained from the atomistic simulations in bulk water and at the cyclohexane water interface, respectively. Using Boltzmann inversion, the B_{AT} was converted to a dihedral potential, for use in the coarse grained simulations. B_{CG} and I_{CG} are the corresponding probability densities obtained from the coarse grained simulations. There is a good agreement between coarse grained and atomistic simulations in bulk water, but the coarse grained simulations failed to reproduce I_{AT} at the cyclohexane/water interface.

4   Conclusions

The MARTINI model for simulations of proteins or peptides was tested by comparing the side-chain partitioning free energies to the Wimley-White scale. The POPC/water partitioning free energies for amino acid P and F are too hydrophobic and will be modified in future versions of MARTINI.

The dihedral potentials derived from the atomistic simulations in bulk water and in turn used in coarse grained simulations failed to reproduce the atomistic behavior at cyclohexane/water interface. This indicates that such potentials are not only specific to amino acid sequence but also to the environment in which the peptide was simulated.

Acknowledgments

References

1. Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries, *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*, J. Phys. Chem. B., **111**, no. 27, 7812–7824, 2007.
2. Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink, *The MARTINI Coarse-Grained Force Field: Extension to Proteins*, J. Chem. Theory Comput., **4**, no. 5, 819–834, 2008.
3. Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput., **4**, no. 3, 435–447, 2008.
4. Michael R. Shirts and John D. Chodera, *Statistically optimal analysis of samples from multiple equilibrium states*, J. Chem. Phys., **129**, no. 12, 124105, 2008.
5. Gurpreet Singh and D. Peter Tieleman, *Using the Wimley-White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model*, J. Chem. Theory Comput., ASAP.

Unfolding Pathways and the Free Energy Landscape of a Single Stranded DNA i-motif

Jens Smiatek[1], Dongsheng Liu[2], and Andreas Heuer[1]

[1] Institute of Physical Chemistry, Westfälische Wilhelms-Universität Münster
48149 Münster, Germany
E-mail: {jens.smiatek, andheuer}@uni-muenster.de

[2] Department of Chemistry, Tsinghua University, Beijing 100084, China
E-mail: liudongsheng@mail.tsinghua.edu.cn

We present Molecular Dynamics simulations of a single stranded DNA i-motif in explicit solvent. Our results indicate that the native structure in non-acidic solution at 300 K completely vanishes on a time scale up to 10 ns. By a comparison of high temperature and biasing potential simulations two unfolding mechanisms can be identified where one pathway is characterized as entropically more favorable. The variations of the entropy can be explained by strong water ordering effects for several structures along the pathways. Finally the lowest free energy configurations belonging to distinct hairpin conformations are indicated in good agreement to experimental results.

1 Introduction

The appearance of non Watson-Crick like structures in DNA has been reported two decades ago[1]. Since this time a lot of effort has been spent to investigate these conformations and possible applications in detail[2]. A prominent representative is the i-motif which is formed in cytosine (C) rich strands of DNA[2].

The stabilizing mechanism for these at a first glance fragile structures is realized by a proton mediated cytosine binding between different strands or regions of the sequence resulting in a stable $C\text{-}CH^+$ pairing[2]. Due to an acidic environment, this is achieved by hemi-protonated cytosines which mimic an ordinary C-G binding as it is present in double helix DNA. Hence it becomes clear that these structures are only occurring at slightly acidic to neutral conditions resulting in pH values from 4.8 to 7.0[1,2]. I-motifs show a remarkable stability and have been found as tetrameric and dimeric complexes although their existence has also been proven for single stranded DNA[2].

The application of this special configuration in modern biotechnology has experienced an enormous growth over the last years[3]. Since the i-motif becomes unstable at pH values larger than 7, a systematic decrease and increase of protons in the solution by changing the pH value results in a reversible folding and unfolding mechanism. It has been shown that this process occurs on a timescale of seconds[3].

Technological applications for this mechanism are given by molecular nanomachines, switchable nanocontainers, pH sensors to detect the pH value inside living cells, building materials for logic gate devices and sensors for distinguishing single walled and multi-walled carbon nanotube systems[3]. Regarding these examples it becomes clear that a detailed investigation of the unfolding pathway of the i-motif is of prior importance.

We present the results of Molecular Dynamics simulations concerning the unfolding mechanism of a maximum unstable single stranded DNA i-motif structure without hemi-

protonated cytosines. Our results indicate a fast initial decay of the i-motif leading to hairpin structures on a timescale up to 10 ns which dominate the unfolded regime in contrast to a fully extended strand. By distinct investigation of the unfolding pathways, two main mechanisms can be identified which significantly differ in their entropic properties. We are able to separate the contributions of the solvent explicitly to determine the influence on the unfolding pathways.

2 Methods

We have performed our Molecular Dynamics simulations of the i-motif in explicit TIP3P solvent at 300 K by the GROMACS software package[4] with the ffAmber03 force field[5]. The single DNA strand consists of 22 nucleic acid bases given by the sequence $5' - CCC - [TAA - CCC]_3 - T - 3'$ where T, A and C denote thymine, adenine and cytosine. The cubic simulation box with periodic boundary conditions has a dimension of $(5.41 \times 5.41 \times 5.41)$ nm filled with 5495 TIP3P water molecules. For a detailed investigation of the unfolding mechanism we conducted five 300 K simulations each with 10 ns duration to calculate the average values.

The calculation of the free energy landscape has been performed by the metadynamics method presented in Refs. 6, 7. The final free energy landscapes have been refined by histogram reweighting of 15 biased simulations of 10 ns at 300 K by the method introduced in Ref. 7.

3 Results

Kinetic investigations concerning the decay of the hydrogen bonds indicate a vanishing of the i-motif on a timescale of 10 ns until a stable hairpin structure occurs[8,9].

To investigate the fully accessible phase space, we applied the metadynamics technique in which a history dependent biasing potential is applied to the molecule which helps to overcome energetic barriers. A demonstration of the decrease of local interactions between the nucleobases is given by the distance matrices and further occurring conformations at later simulation times presented in Fig. 1. It is obvious that the i-motif (Panel a) represents a well-defined structure with many local interactions even for long distances along the backbone. Two further structures (Panels b and c) differ in their nearest neighbor interactions. The structure of (b) has been also derived in the 300 K unbiased simulation run after 7 ns and the structure of (c) is a fully planar hairpin structure which indicates cross-like interactions. Local interactions for all nucleobases between C1-A12 can be observed in (b) whereas (c) indicates interactions between the opposite sides of the strand. In addition it has to be mentioned that we also have observed a fully extended configuration in our simulations (Panel d). This conformation is characterized by the neglect of the cross-like structure shown in Panel (c). It can be concluded that only local nearest neighbor interactions along the backbone are present with short distances between the nucleobases.

The unfolding pathways with the corresponding thermodynamic variables are shown in Fig. 2. The free energy values have been calculated by a metadynamics variant[7] and the corresponding two unfolding mechanisms have been studied by a combination of biased and high temperature simulations[9]. Unfolding pathway 1 is energetically more stable one
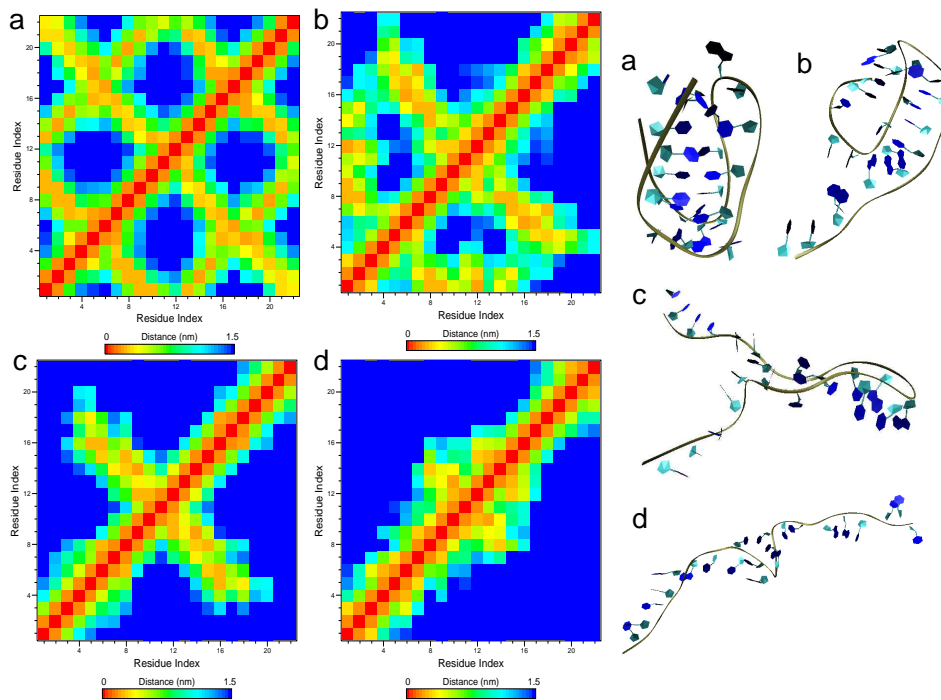
Figure 1. Distance matrices for the 22 nucleobases of the i-motif (left) and three other conformations. Panel a) shows the initial i-motif. Panel b) and c) present the result s for two hairpin structures and Panel d) is related to a fully extended strand.
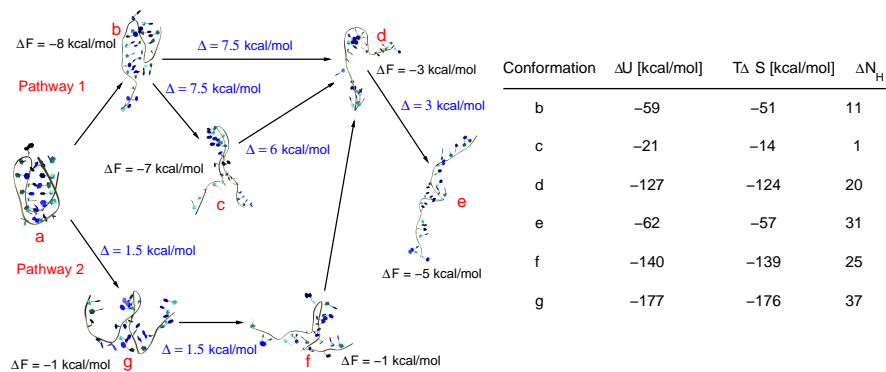

Conformation	ΔU [kcal/mol]	TΔ S [kcal/mol]	ΔN$_H$
b	−59	−51	11
c	−21	−14	1
d	−127	−124	20
e	−62	−57	31
f	−140	−139	25
g	−177	−176	37

Figure 2. Unfolding pathways of the DNA i-motif with the corresponding free energy values ΔF, barriers Δ and denoted conformations (left side), The thermodynamic data concerning the total energy ΔU, the entropic energy $T\Delta S$ and the number of hydrogen bonds Δn_H relative to the i-motif are shown on the right side.

with the global minimum conformations b) and c). It is obvious that the fully extended strand is a local minimum compared to the hairpin structures of b) amd c). Large ener-

getic barriers prevent the molecule from full unfolding. Pathway 2 is energetically less favourable due to higher free energy minima. Furthermore it is obvious that the preferred first unfolding pathway is highly entropically driven, due to the fact that a smaller decrease of entropic contributions can be observed. The decrease of entropic energy can be mainly related due to a significant ordering of the water molecules as it can be seen by the massive increase of hydrogen bonds for structures d)-g). This is mainly caused by an opening of the strands such that the water molecules are able to interact with the nucleobases directly.

In conclusion we have shown that the unfolding of a single stranded i-motif is mainly driven by entropic contributions from the solvent. The final free energy minima are given by distinct hairpin structures in contrast to a fully extended strand. Circular dichroism spectrapolarimetry results have been also indicated that the numerical data are in good agreement to the experimental results[8].

Acknowledgments

References

1. K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, 1993, **363**, 561.
2. M. Gueron and J.-L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326.
3. H. Liu and D. Liu, *Chem. Comm.*, 2009, **19**, 2625.
4. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435.
5. E. J. Sorin and V. S. Pande, *Biophys. J.*, 2005, **88**, 2472.
6. A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562.
7. J. Smiatek and A. Heuer, *J. Comp. Chem.*, 2011, **32**, 2084.
8. J. Smiatek, C. Chen, D. Liu and A. Heuer, *arXiv:1105.3894v1* [physics.bio-ph], submitted (2011).
9. J. Smiatek, D. Liu and A. Heuer, *arXiv:1103.5932v1* [physics.bio-ph], submitted (2011).

The Relationship between Ca^{2+}-Affinity and Shielding of Bulk Water in the Ca^{2+}-Pump from Molecular Dynamics Simulation

Yuji Sugita[1], Mitsunori Ikeguchi[2], and Chikashi Toyoshima[3]

[1] Theoretical Biochemistry Laboratory, RIKEN Advanced Science Institute,
2-1 Hirosawa Wako, Saitama 351-0198, Japan;
Computational Biophysics Research Team, RIKEN Advanced Institute for Computational Science,
7-1-26 Minatojimaminamimachi, Chuo-ku, Kobe 650-0047, Japan;
Laboratory for Biomolecular Function Simulation, RIKEN Quantitative Biology Center,
7-1-26 Minatojimaminamimachi, Chuo-ku, Kobe 650-0047, Japan
E-mail: sugita@riken.jp

[2] Graduate School of Integrated Science, Yokohama City University,
1-7-29 Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan
E-mail: ike@tsurumi.yokohama-cu.ac.jp

[3] Institute for Molecular and Cellular Biosciences, The University of Tokyo,
1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
E-mail: ct@iam.u-tokyo.ac.jp

The sarcoplasmic reticulum Ca^{2+}-ATPase transports two Ca^{2+} per ATP hydrolyzed from the cytoplasm to the lumen against a large concentration gradient. During transport, the pump alters the affinity and accessibility for Ca^{2+} by rearrangements of transmembrane helices. In this study, all-atom molecular dynamics simulations were performed for wild type Ca^{2+}-ATPase in the Ca^{2+}-bound form and the Gln mutants of Glu771 and Glu908. Both of them contribute only one carboxyl oxygen to site I Ca^{2+}, but only Glu771Gln completely looses the Ca^{2+}-binding ability. The simulations show that: (i) For Glu771Gln, but not Glu908Gln, coordination of Ca^{2+} was critically disrupted. (ii) Coordination broke at site II first, although Glu771 and Glu908 only contribute to site I. (iii) A water molecule bound to site I Ca^{2+} and hydrogen bonded to Glu771 in wild type, drastically changed the coordination of Ca^{2+} in the mutant. (iv) Water molecules flooded the binding sites from the lumenal side. (v) The side chain conformation of Ile775, located at the head of a hydrophobic cluster near the lumenal surface, appears critical for keeping out bulk water. Thus the simulations highlight the importance of the water molecule bound to site I Ca^{2+} and point to a strong relationship between Ca^{2+}-coordination and shielding of bulk water, providing insights into the mechanism of gating of ion pathways in cation pumps.

1 Introduction

Ca^{2+}-ATPase of skeletal muscle sarcoplasmic reticulum (SERCA1a) is an integral membrane protein that transports two Ca^{2+} from the cytoplasm into the SR lumen per ATP hydrolyzed against a large concentration gradient across the membrane. Classical E1/E2 theory postulates that the Ca^{2+}-ATPase achieves this by changing the affinity and accessibility of the transmembrane Ca^{2+}-binding sites[1]. That is, the binding sites have high affinity and face the cytoplasm in E1, and have low affinity and face the lumen of SR in E2. The crystal structures of SERCA1a in different states pertinent to the active transport cycle have largely corroborated[2].

SERCA1a consists of three cytoplasmic domains designated as A (actuator), N (nucleotide-binding), P (phosphorylation), and 10 transmembrane helices (M1-M10). Four transmembrane helices (M4-M6, and M8) constitute the two Ca^{2+}-binding sites (I and II) of different characteristics, although Ca^{2+} is coordinated in both sites by 7 oxygen atoms. In site I, all oxygen atoms come from the side chains of residues on 3 helices (M5, M6, and M8) and two water molecules. Asn768 and Glu771 (M5), Thr799 and Asp800 (M6), and Glu908 (M8), all contribute only one oxygen atom to site I Ca^{2+}. In contrast, site II is formed almost on the M4 helix, with contributions by 3 main chain carbonyls (Val304, Ala305, and Ile307 on M4), as well as 4 side chain oxygen atoms: two from Glu309 (M4), one each from Asp800 and Asn796 (M6). The binding geometry in site II is reminiscent of the EF-hand. Asp800 on M6 is the only residue that coordinates both Ca^{2+}. The two sites are located side-by-side but the binding process is single file, with site I Ca^{2+} being the first Ca^{2+} to bind. It has been demonstrated that the removal of only one coordinating oxygen atom from site I completely abolishes Ca^{2+}-binding, yet removal of even 3 oxygen atoms from site II leaves 50% binding[3].

Here we focus on two glutamates in site I: Glu771 and Glu908. Both provide only one carboxyl oxygen for coordination. The Glu908Gln mutation somewhat decreases the affinity for Ca^{2+}, but Glu771Gln is much more deleterious[3]. Why this difference arises is the subject of this study. Simulations demonstrated that Glu908 is likely protonated even when it coordinates site I Ca^{2+}[4], and therefore the small effect of the Glu908Gln mutation is understandable. To explore why Glu771Gln is so deleterious, we apply all-atom molecular dynamics (MD) simulations to the E1·2Ca^{2+} state. In this state, Ca^{2+}-ATPase could be regarded as a high-affinity Ca^{2+}-binding protein. However, the ATPase is special in that it needs entry and exit pathways for both Ca^{2+} and water, and, being a pump, the access needs to be opened and closed, i.e. a control gate needs to be in place. It might appear that the regulation of affinity and that of water accessibility are two different matters. The simulations described here, however, demonstrate that substitution of one oxygen atom with nitrogen in the Glu771 carboxyl critically disrupts both.

2   Simulation Methods

All-atom MD simulations of SR Ca^{2+}-ATPase with two bound Ca^{2+}, explicit solvent, and phospholipids were carried out using MARBLE software package[6]. CHARMM27 force field parameters for proteins[7], phospholipids, and ions, except for Ca^{2+} were used. TIP3P model was used for water molecules. The starting structures for wild-type Ca^{2+}-ATPase bound with two Ca^{2+} were taken from a crystal structure of the enzyme (PDB ID: 1su4). Glu58 and Glu908 were treated as protonated[4]. The starting structures for Glu771Gln and Glu908Gln were made by substituting the Glu in the wild-type structure with Gln keeping the side chain conformation unchanged. Detailed procedures for setting up a full simulation system including a Ca^{2+}-ATPase, 473 DOPC phospholipids, two bound Ca^{2+}, 150 mM salt solution were described previously[4,5].

3   Results and Discussion

In this study, we performed three 10-ns MD simulations for each of the wild type (WT1, WT2, and WT3), Glu771Gln (E771Q1, E771Q2, and E771Q3), and Glu908Gln mutants

(E908Q1, E908Q2, and E908Q3). The root mean square deviations (RMSD) of Cα atoms relative to the crystal structure exceeded 4 in all the simulations. This was due to large rigid body movements of the N and A domains, as the individual domains did not change much (RMSD $\leq$ 2.0). The transmembrane domain was particularly stable (RMSD of around 1.0 for wild type and $\leq$1.5 for the mutants), suggesting that the bound Ca^{2+} ions tie the transmembrane helices together.

In three simulations for wild type, the configuration of the Ca^{2+}-binding sites remained stable with fluctuations between two major forms. In one of the major forms, Ca^{2+} coordination was almost identical to that in the crystal structure, where one of the carboxyl oxygen in Glu771 coordinated site I Ca^{2+} and the other formed a hydrogen bond with a water. In the other stable structure, the Glu771 carboxyl made a bidentate coordination, replacing the Glu908 carboxyl, where instead formed a hydrogen bond with the other water molecule coordinating site I Ca^{2+}. The seven coordination of Ca^{2+} was, thus, maintained in both forms. In the crystal structure, a few water molecules are located outside of Glu309 and in the simulations, they exchanged rapidly with other water molecules in the cytoplasm. A clear water path was observed between the M1 and M2 helices, leading to the Glul309. On the luminal side, a hydrophobic cluster involving Val271(M3), Leu302(M4), Ile775(M5), and Leu787 and Leu792(M6) blocked access of bulk water (Fig. 1). No water molecules from outside the membrane penetrated into the binding sites.
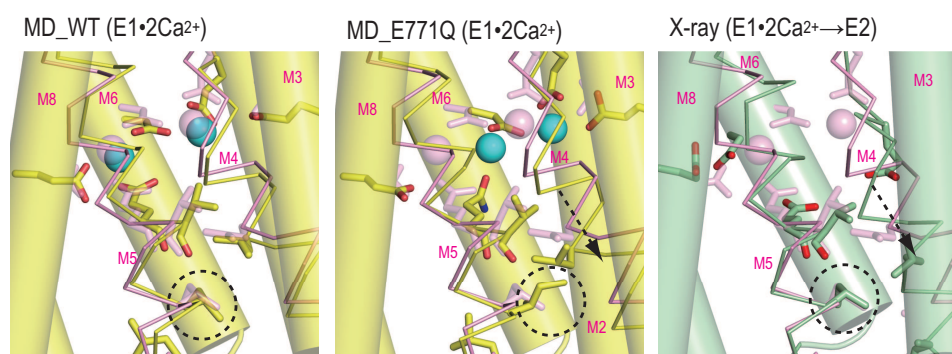
Figure 1. Structural changes in the Ca^{2+}-binding sites. Viewed along the membrane plane. Superimposed on the atomic model of the crystal structure in E1·2Ca^{2+} (pink) are: (A) WT1 at 10 ns (atom color); (B) E771Q3 at 10 ns (atom color); (C) crystal structure in E2 (green). The M3, M6, and M8 helices are represented with cylinders, and M4 and M5 with sticks. The arrows in B and C indicate the movements of the M4 helix relative to the rest of the transmembrane helices. Dashed circles locate the Ile775 side chain.

Coordination of Ca^{2+} in the simulations for Glu908Gln was less stable than that for wild type, whereas, in all three simulations for Glu771Gln, Ca^{2+} coordination was severely altered. These results are consistent with mutagenesis data. In the simulations for Glu771Gln, water molecules were introduced to the binding site from the luminal side, surpassing the hydrophobic cluster. Fig. 1 shows that in the simulations for Glu771Gln, the Ile775 side chain can rotate and relax the hydrophobic cluster involving Val271(M3), Leu302(M4), Leu787(M6), Leu792(M6), and Ile775(M5). Since Ile775 is located just one turn below Glu771 on M5, the mutation effect is transferred to the packing of hydrophobic

residues, resulting in large conformational movements in multiple transmembrane helices (Fig. 1). Thus, the simulations show that one of the reasons why the Glu771Gln mutation is deleterious is that it allows introduction of bulk water into the binding sites.

Acknowledgments

References

1. (a) RW. Albers, Annu.Rev.Biochem. **36**, 727, 1967. (b) RL. Post, S. Kume, C. Hegyvary, J.Biol.Chem. **247**, 6530, 1972.
2. (a) C. Toyoshima, M. Nakasako, H. Nomura and H. Ogawa, Nature **405**, 647, 2000. (b) C. Toyoshima, H. Nomura, Nature **418**, 605, 2002. (c) C. Toyoshima, T. Mizutani, Nature **430**, 529, 2004. (d) C. Toyoshima, H. Nomura, T. Tsuda, Nature **432**, 361, 2004 (e) C. Olesen, T. Sorensen, RC. Nielsen, J. Moller, P. Nissen, Science **306**, 2251, 2004. (f) K. Obara et al., Proc.Natl.Acad.Sci.USA **102**, 14489, 2005. (g) A. Jensen, T. Sorensen, C. Olesen, J. Moller, P. Nissen, EMBO J. **25**, 2305, 2006. (h) M. Takahashi, Y. Kondou, C. Toyoshima, Proc.Natl.Acad.Sci.USA **104**, 5800, 2007. (i) C. Toyoshima, Y. Norimatsu, S. Iwasawa, T. Tsuda, H. Ogawa, Proc.Natl.Acad.Sci.USA **104**, 19831, 2007. (j) C. Olesen et al., Nature **450**, 1036, 2007.
3. ZS. Zhang, DE. Lewis, C. Strock, G. Inesi, Biochemistry **39**, 8758, 2000.
4. Y. Sugita, N. Miyashita, M. Ikeguchi, A. Kidera, C. Toyoshima, J.Am.Chem.Soc. **127**, 6150, 2005.
5. Y. Sugita, M. Ikeguchi, C. Toyoshima, Proc.Natl.Acad.Sci.USA **107**, 21465, 2010.
6. M. Ikeguchi, J.Comp.Chem. **25**, 529, 2004.
7. AD. Mackerell et al., J.Phys.Chem.B **102**, 3586, 1998.

Iterative Computational and Experimental Drug Design Studies of the Complement Inhibitor Compstatin

Phanourios Tamamis[1,2,3]**, Aliana López de Victoria**[2]**, Ronald D. Gorham Jr.**[3]**,**
Meghan Bellows-Peterson[3]**, Christodoulos A. Floudas**[3]**,**
Dimitrios Morikis[2]**, and Georgios Archontis**[1]

[1] Department of Physics, University of Cyprus, PO 20537, Nicosia CY1678, Cyprus

[2] Department of Bioengineering, University of California, Riverside, California 92521, USA

[3] Department of Chemical and Biological Engineering, Princeton University
Princeton, New Jersey 08544, USA
E-mail: tamamis@ucy.ac.cy

The development of drugs regulating complement system activation is of profound medical interest. Compstatin, a 13-residue peptide, binds protein C3 and inhibits complement activation in primate mammals, but is inactive against non-primates. Undoubtedly, the inhibition of non-primate C3 is vital to drug development, since it would allow testing disease models in non-primates. Recently, our molecular dynamics (MD) simulations have elucidated the species specificity of Compstatin. In non-primate simulations, the protein underwent reproducible conformational changes, weakening specific interactions and reducing the complex stability. Subsequently, we aimed at tackling the species specificity using two directions. In the first direction, we *in silico* designed a novel 'transgenic' mouse C3 protein by incorporating specific human-like mutations in mouse C3, and proved that the affinity for Compstatin in the simulations becomes 'human-like'. In the second direction, we conducted simulations of novel Compstatin modified analogs in complex with human and rat/mouse C3. The simulations reveal candidate dual-specificity and high potency human inhibitors, in line with recent experiments.

1   Introduction

Several pathological conditions or complement related diseases, such as macular degeneration and rejection of xenotransplantation can be caused or aggravated by excessive, inappropriate activation of the complement system which constitutes the first line of defense against foreign pathogens. Hence, controlling the complement activation via therapeutic drug-based modulation of the key components of the complement system is of utmost importance (reviewed in Ref. 1).

Compstatin, a 13-residue cyclic peptide, is a promising potent and selective inhibitor of the complement system. It binds directly to the key complement protein C3 and demonstrates clinical potential in a series of experimental model systems. The sequence of the parental peptide is I[CVVQDWGHHRC]T-NH$_2$ and is maintained in a cyclic conformation by a disulfide bridge among cysteins 2 and 12, denoted by brackets. Since its discovery, multiple multidisciplinary biochemical, biophysical experimental and computational studies aimed at the optimization of compstatin binding affinity for C3 and on the elucidation of the key structural-energetic features responsible for its inhibitory potency[1].

2 Structure of Compstatin-C3c Complex

In 2007, B.J.C. Janssen *et al.* revealed the X-ray crystal structure of the most active compstatin analog Ac-Val4Trp/His9Ala (W4A9) comprising natural amino acids, in complex with C3c (a major proteolytic fragment of C3) and showed that the exact binding site of compstatin on C3 the macroglobulin (MG) domains 4 and 5[2]. Based on the structure of the complex and the series of studies performed on compstatin, B.J.C. Janssen *et al.* proposed that compstatin sterically hinders the access of the substrate C3 to the convertase complexes, and consequently it blocks complement activation and amplification.

3 Solution Structure of Compstatin investigated by MD Simulations

Prior to the publication of the X-ray complex structure, we employed multi-ns all-atom/explicit-water simulations of three compstatin analogues with variable activity to investigate the conformational properties of the peptides in aqueous solution[3,4]. We examined the native analogue, the more active mutant W4A9 and the inactive mutant containing one mutation Gln5Gly. The simulations showed that the 5 - 8 segment of free compstatin analogs folds into a 5 - 8 β-turn and the rest of the peptide is mainly disordered, in line with NMR studies[5]. Our results suggested that the critical for activity residue Val3 is involved in reduced hydrophobic contacts and hydrogen bonding interactions in the most active analog and predicted successfully the presence of increased intermolecular Val3-C3 contacts in the W4A9:C3 complex via a hydrophobic cluster involving Val3 and C3 residues, prior to the publication of the experimental X-ray structure[3].

According to the X-ray studies W4A9 binds to C3c in a different conformation compared to the solution conformation: the 5 - 8 β-turn is replaced by a new 8 - 11 β-turn. This result supports the hypothesis that mutations which stabilize the observed bound conformation may yield C3 inhibitors with higher activity[2]. In 2011, new simulations including the CMAP correction showed that the similarity between the simulation conformations and the bound conformation increases with activity, in line with the above hypothesis. Therefore, the results validate this hypothesis[6].

4 Elucidation of the Species Specificity of Compstatin

Compstatin is active against human C3 (including the C3 of several other primate species) but fails to inhibit the activation of C3 from lower mammals. Undoubtedly, the development of effective compstatin inhibitors against non-primate C3 is of profound interest since it can be used to test disease models in non-primate animals (reviewed in Ref. 1).

In 2010, we elucidated this species specificity of compstatin by MD simulations of complexes between the most potent natural compstatin analog and human or rat C3[7]. The results were compared against the experimental X-ray conformation of the human complex, and showed that the human complex simulations preserved faithfully the crystallographic structure. In addition, the human complex simulations provided insights on the relative contributions to stability of specific C3 and compstatin residues. Particularly, residues Ile1 and Cys2 form hydrogen bonds with the side chain of Asn390; the Val3 side-chain participates in a stable hydrophobic cluster with Met346, Pro347, and Leu454; the Trp4 main-chain makes a very stable hydrogen bond with Gly345 CO and a somewhat less stable interaction with the sidechain of Arg456; its side chain packs against the
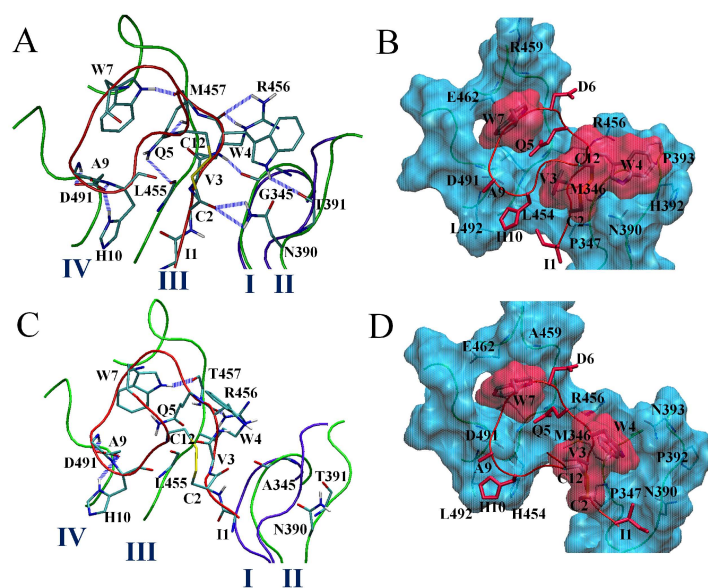
Figure 1. Simulation structures of the compstatin binding site for the human (A, B) and rat (C, D), respectively at the end of MD runs. Important hydrogen bonds are shown in the left panels and non-polar contacts in the right panels. The labels I - IV (in A, C) indicate the four protein sectors with atoms at least within 7 Å. from the ligand. Compstatin is shown in red tubes and sticks. The green tubes show the sectors I-IV. The violet tubes in A, C show the initial conformations of sectors I and II. In B, D protein residues are additionally represented by a cyan color surface, and ligand residues Cys2, Val3, Trp4, Trp7 and Cys12 are represented by a red color surface[7].

Cys2-Cys12 disulfide-bridge on one side and makes nonpolar contacts with Pro393 and the C atom of Gly345 on the other side forming a weak hydrogen bond with Thr391 CO; the Gln5 side-chain makes two intermolecular hydrogen bonds with main-chain groups of Leu455 and Met457; the Trp7 side-chain intercalates between sectors 455 458 and 488 491, making a stable hydrogen bond with Met457 CO and non-polar contacts with Gln5, Met457, Arg459, and Glu462; the main-chain NH groups of Ala9 and His10 form very stable hydrogen bonds with the Asp491 side chain; the His10 sidechain interacts with Asp491 and is near a hydrophobic nucleus formed by Leu454 and Leu492; see Fig. 1 A,B.

In the rat simulations, the protein underwent reproducible conformational changes, eliminating or weakening specific interactions and reducing the complex stability[7]. The displacement of the proximal to the ligand sector 388-393 in the rat complex brought about a loss or weakening of non-polar interactions and hydrogen-bonding between the protein and the ligand, in line with the lack of compstatin activity against rat C3; see Fig. 1 C,D. In general, most protein-ligand interactions become weaker and most intermolecular hydrogen bonds have smaller average occupancies compared to the human complex. The leading changes deal with C3 residues Asn390, Asp491 and Arg456. Specifically, Asn390 loses its hydrogen-bonding interaction with Cys2 due to the displacement of sector 388-393, Asp491 loses a hydrogen-bonding interactions with the His10 mainchain and Arg456 loses polar interactions with Trp4 and non polar interactions with residue Gln5.

5 Computational *de novo* Design and Recent Experimental Studies

Recently, A. López de Victoria *et al.* presented three novel compstatin analogs containing tryptophan mutations at positions 1 and or 13[8]. These analogs were inspired by a physicochemical-reasoning peptide design performed in an earlier work[9]. The design identified tryptophan substitutions at positions 1 and 13 in peptides by predicting C3c binding affinities. The newly experimentally tested analogs preserved the distinct polar and nonpolar surfaces of compstatin and showed altered-improved local interaction capabilities with C3[8]. The results are currently under investigation using MD simulations.

6 Tackling the Species Specificity of Compstatin

We aimed at tackling the speicies specificity towards two different directions. Firstly, we *in silico* designed a novel transgenic mouse C3 protein by incorporating specific human-like mutations in mouse C3, and proved that the affinity for Compstatin in the simulations becomes human-like. Secondly, we conducted simulations of novel Compstatin modified analogs in complex with human and rat/mouse C3. The simulations reveal candidate dual-specificity and high potency human inhibitors, in line with recent experiments[8].

References

1. D. Ricklin, and J.D. Lambris, *Compstatin: a complement inhibitor on its way to clinical application.* Adv. Exp. Med. Biol. **632**, 273-292, 2008.
2. B.J.C. Janssen, E.F. Halff, J.D. Lambris and P. Gros, *Structure of compstatin in complex with complement component C3c reveals a new mechanism of complement inhibition.* J. Biochem. **282**, 29241-29247, 2007.
3. P. Tamamis, S. Skourtis, D. Morikis, J.D. Lambris and G. Archontis, *Conformational analysis of compstatin analogues with molecular dynamics simulations in explicit water.* J. Mol. Graph. Model. **26**, 571-580, 2007.
4. P. Tamamis and G. Archontis, *Secondary structure of Compstatin analogues: Insights from molecular dynamics simulations in Explicit Water.* NIC Series **34**, 133-136, 2006).
5. D. Morikis, N. Assa-Munt, A. Sahu, J.D. Lambris, *Solution structure of compstatin, a potent complement inhibitor.* Protein Sci. **7**, 619-627, 1998.
6. P. Tamamis and G. Archontis, *Solution conformational properties of the potential therapeutic complement inhibitor compstatin and selected analogs, investigated by MD simulations in implicit- and explicit-water.* Biomed. Eng. Res. **1**, 14-21, 2011.
7. P. Tamamis, D. Morikis, C.A. Floudas and G. Archontis, *Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations.* Proteins: Structure, Function, and Bioinformatics **5278**, 2655-2667, 2010.
8. A. López de Victoria, R.D. Gorham Jr., M.L. Bellows, J. Ling, D.D. Lo, C.A. Floudas and D. Morikis, *A new generation of potent complement inhibitors of the compstatin family.* Chem. Biol. Drug. Des. **77**, 431-440, 2011.
9. M.L. Bellows, H.K. Fung, M.S. Taylor, C.A. Floudas, A. López de Victoria and D. Morikis, *New compstatin variants through two de novo protein design frameworks.* Biophys. J. **98**, 2337-2346, 2010.

Multiscale Approach to Thermodynamics and Dynamics of a β-Hairpin Folding

Jacek Wabik, Dominik Gront, Sebastian Kmiecik, and Andrzej Koliński

Department of Chemistry, University of Warsaw, ul. Pasteura 1, 02-093 Warsaw, Poland
E-mail: {jwabik, dgront, sekmi, kolinski}@chem.uw.edu.pl

Formation of the β-hairpin is the first step along the 2GB1 protein folding pathway. This β-structure is one of the nuclei during this process and controls the rate of the whole protein folding. We present an attempt to improve the Replica Exchange Molecular Dynamics (REMD) by utilising the output structures from the coarse-grained Monte Carlo dynamics as the input for the all-atom REMD. This approach enables effective sampling and can be helpful in elucidating the mechanisms of β-hairpin folding. Thermodynamics and dynamics is analyzed focusing on the number of native contacts during simulations. The energy landscape is analyzed by means of the Histogram Method.

1 Introduction

CABS[1] is a mesoscopic CA-CB side-chain protein model which is used in protein structure prediction and validated during Critical Assessment of Protein Structure Prediction (CASP)[2] experiment. Our results indicate that output β-hairpin structures can be very similar to native structure but they have wrong arrangement of the hydrogen bonds. It can be a problem in some applications eg. docking biologically active substances. Formation of the C-terminal β-hairpin is the first step of the 2GB1 peptide folding[2] so it is important to simulate this process. There were many attempts to elucidate the mechanism of β-hairpin formation[3–5]. There are three main ways how the β-hairpin can fold[6]. The first one is the "zipper mechanism"[3]. In this scenario the β structures are formed and stabilized through the hydrogen bond system. According to second mechanism the first step of folding is the hydrophobic collapse of four residues: Trp3, Tyr5, Phe12, and Val14[4,6]. This "core" stabilizes shape of the β-hairpin from the very beginning facilitating formation of hydrogen bonds. The third mechanism was proposed by Felts et al.[5]. They postulated that the zipping of hydrogen bonds and hydrophobic collapse are the simultaneous events.

2 Methods

We used Replica Exchange Molecular Dynamics method with the replica exchange attempted every 4ps. Simulations and most of the analysis have been conducted using GROMACS package[7]. Bioshell[8,9] and do_dssp[10] programs were also employed. Time of the simulation was 150ns per replica. Ten temperature replicas have been distributed in the range of 285K-325K using algorithm of Patricksson and der Spoel[11]. OPLSAA forcefield and tip4p model for the water were used. The equations of motion were integrated using a leap-frog algorithm with a time step of 2 fs. The non-bonded electrostatic interactions were computed using particle-mesh Ewald method and van der Waals interactions using a simple cut-off. Starting conformations to the all-atom MD simulations were selected from

the β-hairpin dynamics simulation by the CABS model. Beta-hairpin CABS dynamics was performed without using any knowledge about the experimental structure (except the weak bias towards the native secondary structure), starting from random conformations. Resulted CABS trajectory exhibited multiple transitions between near native conformations (RMSD around 1 Angstroms, from which the best ones had 0.7 Angstroms) and fully unfolded (most of the fully unfolded conformers had their RMSD around 5 Angstroms). From the CABS trajectory, we have randomly selected 10 conformations, spanning the most frequent resolutions with RMSD to the native from 5 to 0.7 Angstroms. The selected conformations were subjected to the two-step rebuilding procedure, from C-alpha trace to the backbone atoms by the BBQ method[12], and side chains rebuilding by the SCWRL[13] tool. The worst structures start REMD simulation at high temperatures and the best ones at low temperatures. Input structures were minimized with the steepest decent method and equilibrated for 200ps at constant pressure (1013 hPa) and temperature (285K-325K).

3 Results

Altough we use various structures at the beginning, we obtain the average probabilities of the exchange almost equal for every replica (in range of 14%-17%). One can also notice good overlapping of the histograms of the potential energy for various temperatures, which can indicate that the exchange process was efficient.

4 Refinement of the Structures

For four replicas we have obtained the stable β-hairpin structures with ca. 5 of 7 native hydrogen bonds. For one additional replica the structure with ca. three native hydrogen bonds have been obtained. The best conformations from the REMD had CRMSD=0.75Å with the six native hydrogen bonds while the best input conformation had 0.7Å CRMSD with only three native hydrogen bonds. Generally, output conformations with stabilized β structures and native-like turn had CRMSD in the the range of 0.6Å to 2.5Å and we can regard them as the native-like.

5 Mechanism of Folding

We have computed radius of gyration (R_g) of hydrophobic core (Trp3, Tyr5, Phe12, Val14) which should stabilize the β structures[6]. This value for native-like conformations was in range of 5.3Å to 5.7Å. All convergent pseudotrajectories had this value set from the beginning so we cannot check the stabilizing effect of the core. However, there were also conformations with a good value of R_g but without any β structures and they were not not very stable. It indicates that in our simulation folding might not undergo according to hydrophobic collapse. It is interesting that the proper pattern of hydrogen bonds in β structures forms starting from the hairpin ends or sometimes starting from the middle of the structure (Fig. 1).
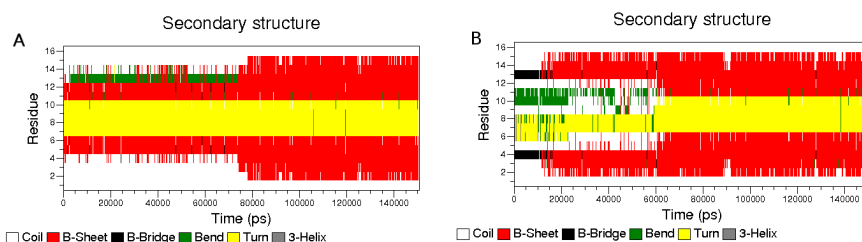
Figure 1. Plots with a secondary structure formed on every residue of β-hairpin for demuxed, continuous trajectories A) pseudotrajectory nr 1 (input CRMSD=1.5Å) B) pseudotrajectory nr 4 (input CRMSD=2.2Å).
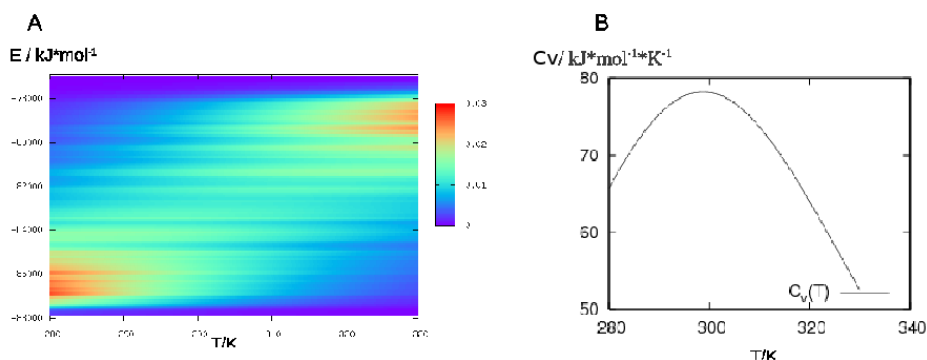

Figure 2. Thermodynamics data A) Plot of density of the states versus T and energy of the system B) Heat capacity versus temperature.

6 Thermodynamic Data

We have computed also some thermodynamic data by the Weighted Histogram Analysis Method (WHAM) technics. Transition temperature corresponding to the maximum of the heat capacity (Fig. 2B) is equal to 298.4K and at this temperature the highest structural mobility and variability was observed. We noticed also that 3- and α-helices constitutes 12% of the population in this temperature, which agrees roughly with the experiment[14] and other simulations. Fig. 2A shows the combined data illustrating the β-hairpin energy landscape and showing cooperative character of the transition with clusters of conformations: native-like and unfolded.

7 Conclusions

We proposed and tested a method that rectifies the conformations generated by coarse-grained Monte Carlo simulation. By REMD we obtained a set of conformations which have proper scheme of native bonds. We also showed that formation of the hydrogen bonds controls the folding mechanism.

Acknowledgments

References

1. Andrzej Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochimica Polonica, **51**, no. 2, 349–371, 2004.
2. Sebastian Kmiecik and Andrzej Kolinski, *Folding pathway of the B1 domain of protein G explored by multiscale modeling*, Biophysical journal, **94**, no. 3, 726–736, Feb. 2008.
3. Victor Muñoz, Eric R. Henry, James Hofrichter, and William A. Eaton, *A statistical mechanical model for β-hairpin kinetics*, Proceedings of the National Academy of Sciences of the United States of America, **95**, no. 11, 5872–5879, May 1998.
4. Angel E. García and Kevin Y. Sanbonmatsu, *Exploring the energy landscape of a β hairpin in explicit solvent*, Proteins, **42**, no. 3, 345–354, 2001.
5. Anthony K. Felts, Yuichi Harano, Emilio Gallicchio, and Ronald M. Levy, *Free energy surfaces of β-hairpin and α-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model*, Proteins, **56**, no. 2, 310–321, Aug. 2004.
6. Phuong H. Nguyen, Gerhard Stock, Emil Mittag, Chin-Kun Hu, and Mai S. Li, *Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study*, Proteins, **61**, no. 4, 795–808, 2005.
7. Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, Journal of Chemical Theory and Computation, **4**, no. 3, 435–447, Mar. 2008.
8. Dominik Gront and Andrzej Kolinski, *Utility library for structural bioinformatics*, Bioinformatics, **24**, no. 4, 584–585, Feb. 2008.
9. Dominik Gront and Andrzej Kolinski, *BioShell–a package of tools for structural biology computations*, Bioinformatics (Oxford, England), **22**, no. 5, 621–622, Mar. 2006.
10. Wolfgang Kabsch and Christian Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*, Biopolymers, **22**, no. 12, 2577–2637, Dec. 1983.
11. Alexandra Patriksson and David Spoel, *A temperature predictor for parallel tempering simulations*, Phys. Chem. Chem. Phys., **10**, no. 15, 2073–2077, 2008.
12. Dominik Gront, Sebastian Kmiecik, and Andrzej Kolinski, *Backbone building from quadrilaterals: A fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates*, J. Comput. Chem., **28**, no. 9, 1593–1597, July 2007.
13. Adrian A. Canutescu, Andrew A. Shelenkov, and Roland L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction*, Protein Science, **12**, no. 9, 2001–2014, Sept. 2003.
14. Victor Munoz and Peggy A. Thompson, *Folding dynamics and mechanism of Beta-hairpin formation.*, Nature, **390**, no. 6656, 196, 1997.

Exploring Energy Landscapes:
From Molecules to Nanodevices

David J. Wales

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom
E-mail: dw34@cam.ac.uk

The first step in exploring the energy landscape for a new system usually involves locating the global potential energy minimum. This article focuses upon basin-hopping global optimisation (Monte Carlo plus minimisation), which has been successfully employed to investigate a remarkably wide range of systems. Examples are chosen to illustrate results for clusters, bulk systems, and a coarse-grained model of helices constructed from colloidal building blocks.

1 Introduction

Methods to explore and sample structure, dynamics and thermodynamics based upon geometry optimisation provide both a conceptual and a computational framework for treating complex potential energy surfaces[1–3]. Coarse-grained treatments of thermodynamics and kinetics can be obtained from stationary points of the potential energy function, $V(\mathbf{X})$. For example, the simplest harmonic superposition approach[4], based only upon local minima, provides an accurate and extremely efficient way to obtain equilibrium thermodynamic properties for low temperature problems involving broken ergodicity. For cases with competing morphologies separated by large barriers compared to $k_B T$, the superposition method can be orders of magnitude faster than parallel tempering[5]. Similarly, calculating transition states and applying unimolecular rate theory for minimum-to-minimum rate constants can provide an approximate view of the overall kinetics for 'rare events'. Here the discrete path sampling approach (DPS) provides a systematic approach to constructing kinetic transition networks, which are analysed using a Markovian assumption[6,7]. No reaction coordinate is required in the DPS approach, just an assignment of products and reactants.

However, before undertaking detailed studies of dynamics and thermodynamics it is probably a good idea to know what is at the bottom of the potential energy landscape, i.e. the global potential energy minimum. This structure is not only the global free energy minimum at low temperature, but, in combination with related defective structures, is also likely to be important in describing the solid phase around the melting temperature[8]. We have also noted that the convergence of parallel tempering calculations can usually be accelerated by starting all the replicas from the global potential energy minimum. Otherwise, we are asking the parallel tempering simulation to perform global optimisation for the lowest temperature replica before equilibration can be achieved. Since dedicated global optimisation algorithms are much faster than parallel tempering for this task, a significant saving of computer time may be possible if the global minimum is identified first. This article will therefore focus on one particular global optimisation technique, introduced by Li and Scheraga[9,10], which has proved successful in a remarkably wide range of applications[1,11,12]. This Monte Carlo plus minimisation approach, or basin-hopping, will be

illustrated for a few selected examples chosen from the Cambridge Cluster Database[13]. Some key references are provided in each case, rather than an exhaustive review, to provide readers with a starting point for accessing the rather extensive global optimisation literature.

2   Methodology

Coupling local minimisation to some sort of search procedure is actually implicit in a number of studies that reported improved global minima for atomic clusters[14–18]. This approach can be viewed as a hypersurface deformation technique, where the potential energy for any configuration, $\mathbf{X}$, becomes the potential energy of the local minimum that our chosen minimisation procedure converges to. More formally, we can write this transformation as

$$\widetilde{V}(\mathbf{X}) = \min\{V(\mathbf{X})\}, \tag{1}$$
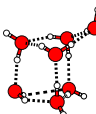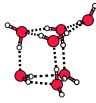
where $\mathbf{X}$ is a $3N$-dimensional vector for a molecule containing N atoms.

We can now consider the transformation of configuration space and the associated search technique separately to define alternative basin-hopping approaches that differ primarily in how steps are taken between minimisations. In the original Monte Carlo plus energy minimization procedure the coordinates were reset to those of the current minimum[9,10]. This approach has generally been found to work more efficiently than allowing the coordinates to vary continuously[18–22], although coordinate resetting violates detailed balance, and must be retained if thermodynamic properties are required, as in the basin-sampling procedure[23]. After proposing a step and minimising the new minimum replaces the old one as the current structure if it satisfies an acceptance condition. Again, various possibilities for this condition have been considered, and the simplest Metropolis test usually works well, i.e. the new minimum with potential energy E_{new} is accepted if it lies below the potential energy of the starting point, E_{old}. If it lies above the starting point then the step is accepted if $\exp[(E_{\text{old}} - E_{\text{new}})/kT]$ is greater than a random number drawn from the interval [0,1]. The temperature associated with the Metropolis test then becomes an important parameter of the procedure, along with the magnitude of the proposed steps. Threshold acceptance and other non-Boltzmann sampling schemes have been found to give acceptable performance in basin-hopping, and procedures that only accept downhill moves have also proved useful[24]. However, if further parameters are introduced then the optimal values are likely to become system specific.

One reason why the minimisation step is crucial to the success of basin-hopping is that much larger steps can be taken in configuration space, since the energy that is considered is the value after minimisation. Hence atoms can pass through each other, and transitions between basins of attraction between different minima can occur anywhere along the boundaries between basins of attraction. Downhill barriers are effectively removed by this procedure, thus accelerating the search of configuration space. However, these observations on their own do not explain why basin-hopping succeeds for multi-funnel landscapes[25] corresponding to different morphologies separated by high barriers. In fact there is another factor working in favour of the transformation, because the occupation probabilities of competing regions of configuration space overlap over a temperature range

where interconversion is still possible[19,26]. Hence the problem of trapping in low, but sub-optimal, minima is partly alleviated.

3 Examples

n	TIP5P				TIP4P		
	Structure	H-bonds	Point group	Energy	Structure	H-bonds	Remin energy
2		1	C_s	-28.3878		1	-28.3878
3		3	C_1	-62.7276		3	-62.7276
4		4	S_4	-118.9554		4	-118.9554
5		5	C_1	-159.5019		5	-159.5019
6		6	S_6	-197.9372		8	-189.5730
7		10	C_1	-242.0671		10	-242.0671
8		12	D_{2d}	-303.1984		12	-301.3378
9		13	C_1	-349.5400		13	-349.5400
10		15	C_1	-399.0432		15	-397.3824
11		16	C_1	-440.5860		17	-427.2306
12		18	C_1	-494.9000		20	-475.4220
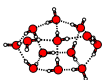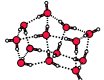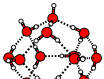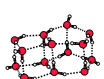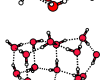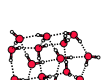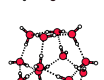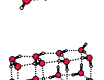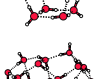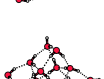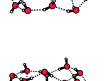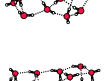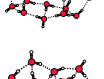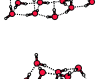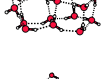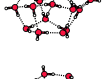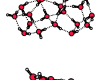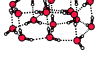
13		19	C_1	-535.4792		21	-524.5022
14		21	C_1	-591.6422		23	-579.0053
15		23	C_1	-631.8994		25	-626.8068
16		24	C_2	-685.2407		28	-656.2767
17		27	C_1	-729.4807		29	-702.7235
18		28	C_1	-780.3213		31	-759.3826
19		31	C_1	-833.3290		31	-819.8277
20		32	C_1	-881.3449		34	-863.1006
21		34	C_1	-935.3950		37	-885.2192

Table 1: Comparison of likely global minima for TIP5P $(H_2O)_N$ clusters, $N = 2$-21, and TIP4P structures[21], including the energies of the latter geometries when relaxed with TIP5P. All energies are in $\mathrm{kJ\,mol^{-1}}$.

3.1 Water Clusters

A vast number of alternative intermolecular potentials have been proposed for water, in accord with the central importance of the liquid phase for biomolecular and atmospheric simulations. It is interesting to compare the global minima for water clusters, $(H_2O)_N$, with different potentials as a function of N to relate the corresponding structures to the predicted behaviour for bulk water. The observed trends may also point to deficiencies in the empirical parameterisation, and provide insight into how this may be improved.

The first basin-hopping study of water clusters[21] provided results for the TIP4P potential[27,28] up to N=21. These predictions have all been confirmed in subsequent studies[29,30], and were compared with a five-site TIP5P model[31] by James et al.[32] in 2005. The results are summarised in Tab. 1 and the structures can be downloaded from the Cambridge Cluster Database[13]. The main difference between the two models is that the negative partial

charge is shifted and split in the TIP5P potential, with the aim of representing the oxygen lone pairs more realistically. When the TIP4P global minima are relaxed with the TIP5P potential $(H_2O)_{13}$ and $(H_2O)_{17}$ lose a single hydrogen-bond, while $(H_2O)_{21}$ loses two. The rest are topologically stable, but start to differ from the TIP5P global minima beginning with $(H_2O)_6$ (Tab. 1). The clearest systematic trends are that TIP4P exaggerates the relative stability of fused cubic structures compared to those containing five- and six-membered rings[33], while TIP5P global minima tend to have fewer hydrogen-bonds than the TIP4P counterparts[32]. These results are in line with previous work, which suggested that suggested tetrahedral arrangements of hydrogen-bonds are favoured for TIP5P[31].

Locating the global minima for water clusters is significantly harder than for most atomic clusters with a comparable number of degrees of freedom[21,30,34]. The added difficulty comes from the coupling of translational (centre of mass) and orientational degrees of freedom. With the molecular centres of mass roughly constant the possible hydrogen-bonding patterns can span a wide range of energy, and a two-stage search procedure is required to locate the global minimum efficiently. Previous basin-hopping schemes have employed blocks of moves involving only centre-of-mass displacements alternating with moves involving only angular displacements[21]. More systematic, but less general approaches (specific to water), have employed prescreening of the hydrogen-bonding patterns to identify the best candidates[35–45]. The effect of the coupling between translation and orientation has been visualised for the $(H_2O)_{20}$ cluster, revealing a hierarchical potential energy landscape, with low-lying minima separated by high barriers[1,7,46]. This structure is associated with multiple relaxation time scales due to the large number of kinetic traps[2,23].

3.2 The Thomson Problem

J. J. Thomson formulated his famous problem in 1904 as a model for atomic structure[48]. The question he posed is to find the lowest energy for N unit charges constrained to a sphere, with potential energy (atomic units)

$$V = \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|},$$ (2)

where the radius is $|\mathbf{r}_\alpha| = 1$ for all particles α. Although this Thomson model did not prove to be useful for analysing the electronic structure of atoms, it has instead provided insight into the structure of a wide variety of systems constrained to have spherical topology. Examples include spherical viruses[49,50], multielectron bubbles in superfluid helium[51,52], 'colloidosomes'[53–55], superconducting films[56,57], lipid rafts deposited on vesicles[58], cell surface layers in prokaryotic organisms[59,60], colloidal silica microspheres[61], and micropatterning of spherical particles[62].

Euler's theorem states that the total disclination charge must be 12 for a triangulated structure defined by a set of particles constrained to a spherical surface. Here the disclination charge is defined as $Q = 6 - C$, where C is the number of neighbours for a particle confined to the surface. The simplest way to satisfy this topological constraint is for 12 particles to form five-coordinate disclinations. A number of global minima for the Thomson problem at smaller sizes do indeed conform to this pattern, but alternative defect structures occur as N increases, providing less strained geometries and lower energies[47,63]. For
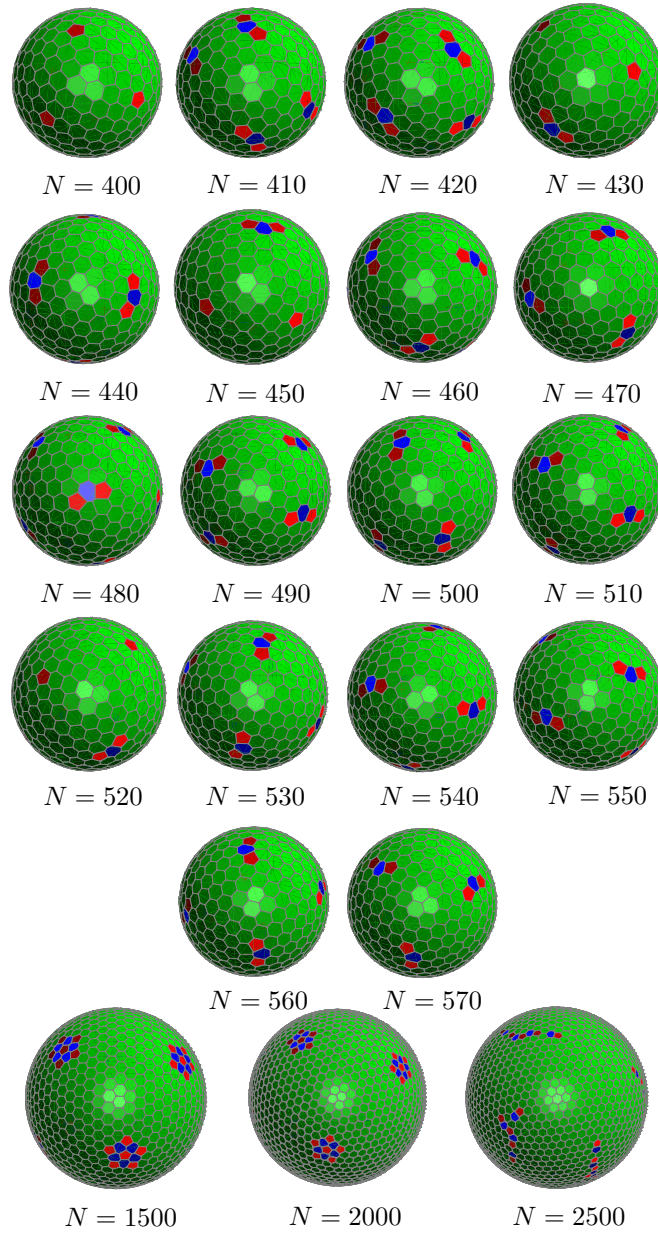
Figure 1. Voronoi representations of the lowest minima located for $N = 400,\ 410,\ldots,\ 570$ and $N = 1500$, 2000 and 2500. The pentagons, hexagons and heptagons are coloured red, green and blue, respectively[47].

example, dislocations consisting of adjacent five-coordinate and seven-coordinate particles[64–66] are observed, and extended dislocations containing a heptagon and two pentagons in the Voronoi representation carry the same unit disclination charge as a single pentagon.

These features are favourable for intermediate sizes[63], in agreement with predictions from an continuum elastic model[67]. As the number of particles increases further we find global minima containing twinned defects, where two heptagons share an edge in the Voronoi construction, with three pentagons on the periphery. Again, the disclination charge is unity. For $N > 400$ twinned defects and embryonic grain boundaries appear to be the most common motifs in the lowest known minima (Fig. 1)[47,63].

Since the number of local minima generally increases exponentially with size[68,69], global optimisation becomes correspondingly more difficult. However, the rate of increase for the Thomson problem is smaller than for systems bound by shorter-range potentials. The long-range Coulomb potential supports fewer local minima, for reasons that are well understood[70-73]. Nevertheless, locating the true global minimum for the Thomson problem with a thousand or more particles can still be challenging[47]. Pursuing these searches may be very rewarding, particularly when remarkable structures appear for particular sizes. Many putative global minima at larger sizes exhibit 'rosette' features in the Voronoi representation, where a central pentagon is surrounded by five heptagons alternating with five additional pentagons (Fig. 1). In fact, there are candidate global minimum structures with icosahedral symmetry, which exhibit twelve such rosettes, thereby achieving the correct disclination charge[47]. These structures were anticipated using explicit construction by Pérez-Garrido and Moore[65], and are referred to as 'pentagonal buttons' in a report that employs continuum elastic theory[74], which cites unpublished work by A. Toomre. Basin-hopping global optimisation has now identified structures with point group I at $N = 1632$ and 1902 as likely global minima. It will be very interesting to see whether these motifs can be identified in any experimental systems in the future.

3.3 A Binary Glass-Former

Predicting crystal structures can also present a formidable problem for global optimisation[81,82]. Some of the issues are common to the protein structure prediction problem[83], while others, such as the size, shape, and content of the supercell, are peculiar to bulk packing. There have been relatively few applications of basin-hopping to crystal structure to date, and this is an area where increased activity is likely in the near future. One notable success was the location of a crystal structure for the popular binary Lennard-Jones (BLJ) system, which some workers in the field believed to have no crystal. BLJ models have proved very popular models in the supercooled liquids and glasses field, since they can be designed to disfavour crystallisation on normal molecular dynamics time scales[84-89,77,90-92]. The system described here corresponds to an 80:20 mixture of A:B atoms with parameters $\sigma_{AA} = 1$, $\sigma_{AB} = 0.8$, $\sigma_{BB} = 0.88$, $\epsilon_{AA} = 1$, $\epsilon_{AB} = 1.5$, and $\epsilon_{BB} = 0.5$ [84]. The Stoddard-Ford scheme was used to ensure that the energy and first derivatives were continuous at the cut-off[93]. A non-phase-separated crystalline state was first reported using basin-hopping[78], and a still lower phase-separated state has been explicitly constructed[94]. The unit cell for the non-phase-separated crystal has space group $I4/mmm$, with a structure analogous to $Ir(UC)_2$.
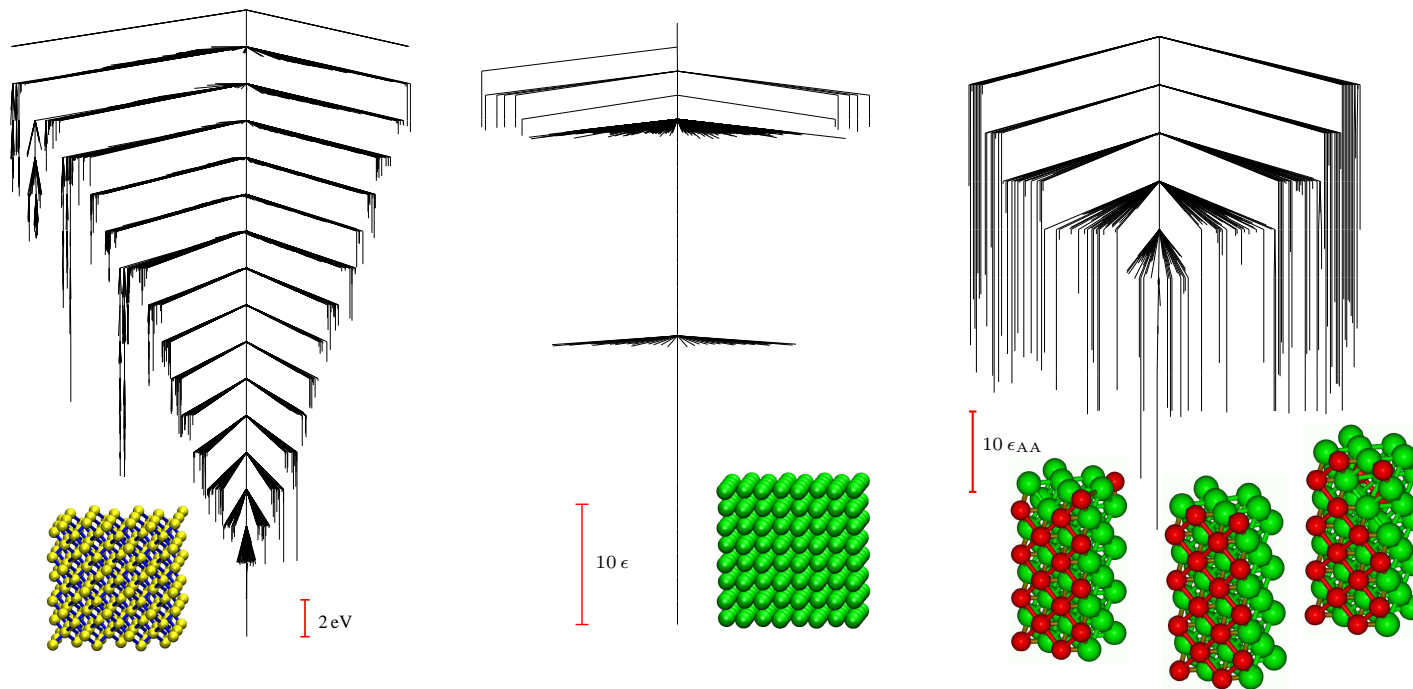
Figure 2. Disconnectivity graphs for three crystalline systems; some of the low-lying minima are illustrated near to their corresponding branches. Left: disconnectivity graph obtained at one atmosphere pressure for a supercell of 216 atoms using a modified Stillinger-Weber silicon potential[75] where the magnitude of the three-body term is increased by 50%[1]. Middle: disconnectivity graph including the lowest 500 minima for the unit density Lennard-Jones[76] crystal with a supercell containing 256 atoms[77]. Right: disconnectivity graph for a binary Lennard-Jones crystal represented by a supercell containing 320 atoms in an 80:20 A:B ratio[78].
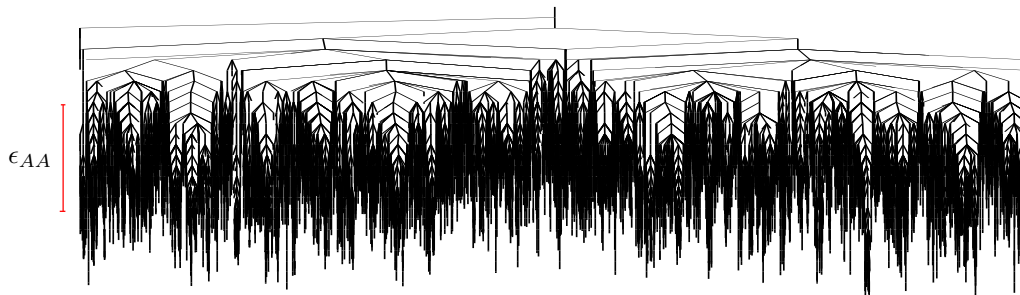
Figure 3. Disconnectivity graph obtained for a BLJ system with 60 atoms in the supercell at a number density of $1.3\sigma_{\mathrm{AA}}^{-3}$ and $k_B T/\epsilon_{\mathrm{AA}} = 0.713$. The minima and transition states correspond to a locally ergodic molecular dynamics trajectory[79] defined using the Mountain and Thirumalai energy fluctuation metric[80].

The additional complexity that arises for the BLJ system is visible even in low-lying regions of the potential energy surface around the crystal. Fig. 2 compares the corresponding disconnectivity graph[46,95] with those obtained for two systems that readily crystallise. The BLJ graph exhibits low-lying minima separated by high barriers, which characterise a 'frustrated' landscape[96,97]. At higher energy, the BLJ landscape exhibits glassy characteristics, with numerous local minima corresponding to amorphous structures separated by barriers that are very large compared to the available thermal energy at relevant temperatures (Fig. 3)[98–100]. This structure is essentially the opposite of that expected for a good 'structure-seeking' system corresponding to minimal frustration[1–3,97,101]. Crystal structure prediction for molecular glass-formers, especially those involving hydrogen-bonding[102–105], is unlikely to become a routine procedure for some time.

3.4 Modelling Mesoscopic Systems

A fully atomistic approach to modelling mesoscopic systems is usually neither practical nor necessary. Often we are interested in generic properties, and understanding the minimum conditions for which particular structures are favourable. For example, a wide range of alternative shells, tubes, spirals and helices can be formed by relatively simple building blocks if we tune the shape of the interactions between them[107–110]. Here we have considered constituent particles as rigid bodies, using angle-axis coordinates to describe orientation[2,111]. (This scheme was also used for the TIP5P water clusters described in Section 3.1.) The Paramonov-Yaliraki potential can be used to define the shape of single-site attraction and repulsion terms[112], with additional sites decorating the rigid body. This representation is efficient enough to provide a framework for analysing how the global potential energy landscapes changes with the parameterisation, which enables us to identify systems capable of self-assembly.

The first step in this process always involves identifying the global minimum. One particular design principle that has recently been employed to construct helical strands is the interplay between competing length scales[106,108]. Asymmetric dipolar dumbbells were used to model experiments involving colloidal building blocks[113]. Each dumbbell involves two Lennard-Jones sites with a point-dipole directed across the axis between them, and
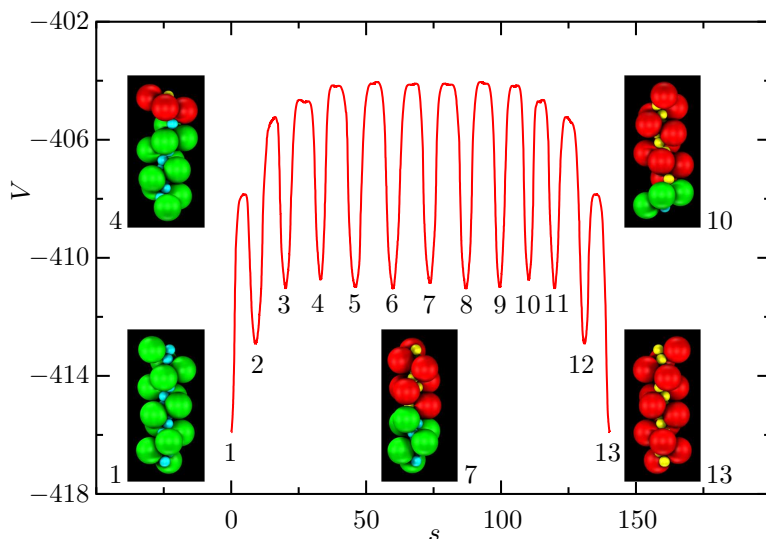
Figure 4. Potential energy, V, as a function of integrated Cartesian path length along the pathway, s for conversion of left- and right-handed helices composed of 13 dipolar dumbbells. Left- and right-handed portions of the helix are coloured red/yellow and green/cyan in local minima illustrated along the profile[106].

the ratio of the diameters of the spherical lobes then defines an asymmetry parameter, α. When an assembly of such particles is simulated in a constant electric field the global minimum becomes a helix above a critical value of α. The competing length scales are the steric interactions and the tendency of the dipoles to align with the applied field. Both left- and right-handed helices then exist, with equal energies, and the pathway between them exhibits some remarkable properties[106]. The energy profile for this path is illustrated in Fig. 4. The mechanism involves migration of the boundary between regions with opposite handedness via a series of transition states. Each of these steps corresponds to an elongation of the chain, so that elongated and compact states of the helix alternate along the path. At the same time, the terminal dumbbells rotate in opposite directions throughout the path. Hence this pathway involves coupled linear and rotary motion, and in ongoing work we are seeking a way to drive the transition using an external energy reservoir.

4   Conclusions

Global optimisation will often be the first calculation to run on a new model system. This approach can result in large savings of computer time and human effort if detailed thermodynamic and kinetic properties are of interest. For instance, it may be necessary to run parallel tempering simulations for hundreds of nanoseconds before escape from a high energy local minimum is achieved, whereas global optimisation will produce lower-lying minima in a few steps. Hence global optimisation runs are likely to be much more efficient for investigating the stability of a structure obtained by construction, or for testing a known

crystal structure with an empirical force field.

Basin-hopping global optimisation, which covers any procedure where steps are taken between local minima, has proved to be a very effective tool. In particular, it is readily transferable between different systems, because very few parameters need to be specified in advance. Since optimal efficiency can only be obtained once we know the correct answer, the ability to obtain good candidates for the global minimum without an extensive search of parameter space is particularly valuable.

The examples illustrated above are simply intended to show some of capabilities of the basin-hopping approach. In the future, such calculations are likely to play a key role in the design of self-assembling systems, where we must verify that the target morphology is indeed favourable under experimental conditions. To determine whether this structure is kinetically accessible generally requires further calculations[2,3,8], but global optimisation can help to guide these searches into regions of parameter space that are likely to be most productive.

References

1. D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge (2003).
2. D. J. Wales, Phil. Trans. Roy. Soc. A **363**, 357 (2005).
3. D. J. Wales, Curr. Op. Struct. Biol. **20**, 3 (2010).
4. B. Strodel and D. J. Wales, Chem. Phys. Lett. **466**, 105 (2008).
5. V. A. Sharapov, D. Meluzzi and V. A. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).
6. D. J. Wales, Mol. Phys. **100**, 3285 (2002).
7. D. J. Wales, Mol. Phys. **102**, 891 (2004).
8. D. J. Wales and T. V. Bogdan, J. Phys. Chem. B **110**, 20765 (2006).
9. Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).
10. Z. Li and H. A. Scheraga, J. Mol. Struct. (Theochem) **179**, 333 (1988).
11. D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).
12. K. Klenin, B. Strodel, D. J. Wales and W. Wenzel, Biochimica et Biophysica Acta **1814**, 977 (2011).
13. D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas and T. F. Middleton, The Cambridge Cluster Database, http://www-wales.ch.cam.ac.uk/CCD.html (2001).
14. G. L. Xue, J. Glob. Opt. **4**, 187 (1991).
15. C. Barrón, S. Gómez and D. Romero, App. Math. Lett. **9**, 75 (1996).
16. D. M. Deaven, N. Tit, J. R. Morris and K. M. Ho, Chem. Phys. Lett. **256**, 195 (1996).
17. C. Barrón, S. Gómez and D. Romero, App. Math. Lett. **10**, 25 (1997).
18. D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).
19. J. P. K. Doye and D. J. Wales, Phys. Rev. Lett. **80**, 1357 (1998).
20. R. P. White and H. R. Mayne, Chem. Phys. Lett. **289**, 463 (1998).
21. D. J. Wales and M. P. Hodges, Chem. Phys. Lett. **286**, 65 (1998).
22. J. P. K. Doye and D. J. Wales, New J. Chem. **22**, 733 (1998).
23. T. V. Bogdan, D. J. Wales and F. Calvo, J. Chem. Phys. **124**, 044102 (2006).
24. R. H. Leary and J. P. K. Doye, Phys. Rev. E **60**, R6320 (1999).
25. J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).

26. J. P. K. Doye, D. J. Wales and M. A. Miller, J. Chem. Phys. **109**, 8143 (1998).
27. W. L. Jorgensen, J. Am. Chem. Soc. **103**, 335 (1981).
28. W. L. Jorgensen, J. Chem. Phys. **77**, 4156 (1982).
29. H. Kabrede and R. Hentschke, J. Phys. Chem. B **107**, 3914 (2003).
30. B. Hartke, Phys. Chem. Chem. Phys. **5**, 275 (2003).
31. M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).
32. T. James, D. J. Wales and J. Hernández-Rojas, Chem. Phys. Lett. **415**, 302 (2005).
33. C. J. Tsai and K. D. Jordan, J. Phys. Chem. **97**, 5208 (1993).
34. B. Hartke, Chem. Phys. **346**, 286 (2008).
35. S. Kazachenko and A. Thakkar, Mol. Phys. **108**, 2187 (2010).
36. D. J. Anick, J. Chem. Phys. **132**, 164311 (2010).
37. P. Bandyopadhyay, Chem. Phys. Lett. **487**, 133 (2010).
38. S. Kazachenko and A. J. Thakkar, Chem. Phys. Lett. **476**, 120 (2009).
39. M. V. Kirov, G. S. Fanourgakis and S. S. Xantheas, Chem. Phys. Lett. **461**, 180 (2008).
40. A. Khan, J. Mol. Struct. **850**, 144 (2008).
41. Q. Shi, S. Kais and J. S. Francisco, J. Phys. Chem. A **109**, 12036 (2005).
42. A. Lenz and L. Ojamae, Phys. Chem. Chem. Phys. **7**, 1905 (2005).
43. J. L. Kuo, J. V. Coe, S. J. Singer, Y. B. Band and L. Ojamae, J. Chem. Phys. **114**, 2527 (2001).
44. M. D. Tissandier, S. J. Singer and J. V. Coe, J. Phys. Chem. A **104**, 752 (2000).
45. S. McDonald, L. Ojamae and S. J. Singer, J. Phys. Chem. A **102**, 2824 (1998).
46. D. J. Wales, M. A. Miller and T. R. Walsh, Nature **394**, 758 (1998).
47. D. J. Wales, H. McKay and E. L. Altschuler, Phys. Rev. B **79**, 224115 (2009).
48. J. Thomson, Philos. Mag. **7**, 237 (1904).
49. D. L. D. Caspar and A. Klug, Cold Spring Harbour, Symp. Quant. Biol. **27**, 1 (1962).
50. C. J. Marzec and L. A. Day, Biophys. J. **65**, 2559 (1993).
51. U. Albrecht and P. Leiderer, J. Low Temp. Phys. **86**, 131 (1992).
52. P. Leiderer, Z. Phys. B **98**, 303 (1995).
53. A. R. Bausch, A. C. M. J. Bowick, A. D. Dinsmore, M. F. Hsu, D. R. Nelson, M. G. Nikolaides, A. Travesset and D. A. Weitz, Science **299**, 1716 (2003).
54. P. Lipowsky, M. J. Bowick, J. H. Meinke, D. R. Nelson and A. R. Bausch, Nat. Mater. **4**, 407 (2005).
55. T. Einert, P. Lipowsky, J. Schilling, M. J. Bowick and A. R. Bausch, Langmuir **21**, 12076 (2005).
56. M. J. W. Dodgson, J. Phys. A **29**, 2499 (1996).
57. M. J. W. Dodgson and M. A. Moore, Phys. Rev. B **55**, 3816 (1997).
58. K. Simons and W. L. C. Vaz, Annu. Rev. Biophys. Biomol. Struct. **33**, 269 (2004).
59. D. Pum, P. Messner and U. B. Sleytr, J. Bacteriol. **173**, 6865 (1991).
60. U. B. Sleytr, M. Sára, D. Pum and B. Schuster, Prog. Surf. Sci. **68**, 231 (2001).
61. Y.-S. Cho, G.-R. Yi, J.-M. Lim, S.-H. Kim, V. N. Manoharan, D. J. Pine and S.-M. Yang, J. Am. Chem. Soc. **127**, 15968 (2005).
62. Y. Masuda, T. Itoh and K. Koumoto, Adv. Mater. **17**, 841 (2005).
63. D. J. Wales and S. Ulker, Phys. Rev. B **74**, 212101 (2006).
64. D. R. Nelson, *Defects and Geometry in Condensed Matter Physics*, Cambridge University Press, Cambridge (2002).

65. A. Pérez-Garrido and M. A. Moore, Phys. Rev. B **60**, 15628 (1999).
66. A. Pérez-Garrido, Phys. Rev. B **62**, 6979 (2000).
67. M. J. Bowick, A. Cacciuto, D. R. Nelson and A. Travesset, Phys. Rev. B **73**, 024115 (2006).
68. F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
69. D. J. Wales and J. P. K. Doye, J. Chem. Phys. **119**, 12409 (2003).
70. J. P. K. Doye, D. J. Wales and R. S. Berry, J. Chem. Phys. **103**, 4234 (1995).
71. J. P. K. Doye and D. J. Wales, Science **271**, 484 (1996).
72. J. P. K. Doye and D. J. Wales, J. Chem. Soc., Faraday Trans. **93**, 4233 (1997).
73. D. J. Wales, Science **293**, 2067 (2001).
74. M. J. Bowick, D. R. Nelson and A. Travesset, Phys. Rev. B **62**, 8738 (2000).
75. F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).
76. J. E. Jones and A. E. Ingham, Proc. R. Soc. A **107**, 636 (1925).
77. T. F. Middleton and D. J. Wales, Phys. Rev. B **64**, 024205 (2001).
78. T. F. Middleton, J. Hernández-Rojas, P. N. Mortenson and D. J. Wales, Phys. Rev. B **64**, 184201 (2001).
79. V. K. de Souza and D. J. Wales, J. Chem. Phys. **123**, 134504 (2005).
80. R. D. Mountain and D. Thirumalai, J. Phys. Chem. **93**, 6975 (1989).
81. Y. A. Arnautova, A. Jagielska, J. Pillardy and H. A. Scheraga, J. Phys. Chem. B **107**, 7143 (2003).
82. J. D. Dunitz and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **101**, 14309 (2004).
83. J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy, J. A. Saunders, K. D. Gibson and H. A. Scheraga, Int. J. Quant. Chem. **77**, 90 (2000).
84. W. Kob and H. C. Andersen, Phys. Rev. Lett. **73**, 1376 (1994).
85. W. Kob and H. C. Andersen, Phys. Rev. E **51**, 4626 (1995).
86. W. Kob and H. C. Andersen, Phys. Rev. E **52**, 4134 (1995).
87. S. Sastry, P. G. Debenedetti and F. H. Stillinger, Nature **393**, 554 (1998).
88. S. Büchner and A. Heuer, Phys. Rev. E **60**, 6507 (1999).
89. F. Sciortino, W. Kob and P. Tartaglia, Phys. Rev. Lett. **83**, 3214 (1999).
90. T. F. Middleton and D. J. Wales, J. Chem. Phys. **118**, 4583 (2003).
91. F. Sciortino and P. Tartaglia, Phys. Rev. Lett. **86**, 107 (2001).
92. S. Sastry, Nature **409**, 164 (2001).
93. S. D. Stoddard and J. Ford, Phys. Rev. A **8**, 1504 (1973).
94. J. R. Fernández and P. Harrowell, Phys. Rev. E **67**, 011403 (2003).
95. O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
96. J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
97. J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, Proteins: Struct., Func. and Gen. **21**, 167 (1995).
98. F. Calvo, T. V. Bogdan, V. K. de Souza and D. J. Wales, J. Chem. Phys. **127**, 044508 (2007).
99. V. K. de Souza and D. J. Wales, J. Chem. Phys. **129**, 164507 (2008).
100. V. K. de Souza and D. J. Wales, J. Chem. Phys. **130**, 194508 (2009).
101. J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
102. G. M. Day, J. Chisholm, N. Shan, S. Motherwell and W. Jones, Crystal Growth and Design **4**, 1327 (2004).

103. G. M. Day, S. Motherwell and W. Jones, Phys. Chem. Chem. Phys. **9**, 1693 (2007).
104. S. M. Woodley and R. Catlow, Nature Materials **7**, 937 (2007).
105. S. Price, Phys. Chem. Chem. Phys. **10**, 1996 (2008).
106. D. Chakrabarti and D. J. Wales, Soft Matter **7**, 2325 (2011).
107. S. N. Fejer and D. J. Wales, Phys. Rev. Lett. **99**, 086106 (2007).
108. D. Chakrabarti, S. N. Fejer and D. J. Wales, Proc. Nat. Acad. Sci. USA **106**, 20164 (2009).
109. S. N. Fejer, D. Chakrabarti and D. J. Wales, ACS Nano **4**, 219 (2010).
110. S. N. Fejer, D. Chakrabarti and D. J. Wales, Soft Matter **7**, 3553 (2011).
111. D. Chakrabarti and D. Wales, Phys. Chem. Chem. Phys. **11**, 1970 (2009).
112. L. Paramonov and S. N. Yaliraki, J. Chem. Phys. **123**, 194111 (2005).
113. D. Zerrouki, J. Baudry, D. Pine, P. Chaikin and J. Bibette, Nature **455**, 380 (2008).

Generalized Correlations in Molecular Evolution: A Critical Assessment

Stephanie Weißgraeber and Kay Hamacher

Bioinformatics & Theoretical Biology, Dept. of Biology, Technische Universität Darmstadt
64287 Darmstadt, Germany
E-mail: {weissgraeber, hamacher}@bio.tu-darmstadt.de

Analyzing coevolutionary relationships in proteins offers the possibility to identify "interaction" partners, e.g., residues, within or between proteins. Previously suggested approaches to quantify such relationships include *statistical coupling analysis*, which determines the degree of coevolution by either performing a perturbation analysis or obtaining empirical, classical correlation measures. Computing the *mutual information* of a protein multiple sequence alignment is a well-established approach to evaluate the correlation between amino acids. Here, we compare this information-theoretical method to the statistical coupling analysis and discuss the underlying concepts.

1 Introduction

Mutations in the genome change the primary structure of a protein and can thereby alter the chemical and physical identity of an amino acid – leading to perturbed interactions in the protein. Since there is a selective pressure to maintain the function of the protein, compensating mutations are favored. Understanding this evolutionary pressure on the molecular phenotype requires the detailed analysis of such coevolutionary relationships[1] within or between proteins.

Using a sufficiently large number of homologous sequences one constructs multiple sequence alignments (MSA) and uses these as input for coevolutionary analysis. This step can be performed, however, using different concepts, whose respective advantages and shortcomings we discuss in this contribution.

2 Statistical Coupling Analysis

Statistical coupling analysis (SCA) is based on the assumption that every position in an MSA whose amino acid distribution differs from the mean is conserved to a certain extent. The mean distribution is derived from histograms of the entries in a database, e.g., of all protein sequences deposited in Swiss-Prot's non-redundant database[2]. In the original formulation of SCA, an empirical evolutionary conservation parameter ΔG^{stat} is defined for site i:

$$\Delta G_i^{\text{stat}} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \tag{1}$$

with kT^* being an arbitrary energy unit, P_i^x the probability of any amino acid x at site i expressed as binomial probability of the observed number of amino acid type x given

its mean frequency derived from the database as described above, and P_{MSA}^x the mean frequency of amino acid x in the MSA under consideration, again written as a binomial probability.

To measure the relationship between two sites i and j, a perturbation of the amino acid frequencies at another site j ($\Delta G_{i|\delta j}^{\text{stat}}$) is simulated. This is done by selecting a subset of the MSA choosing only those sequences in which the most prevalent amino acid of position j can be found. The magnitude of the difference of the statistical energy vectors at position i considering either the full alignment (ΔG_i^{stat}) or the selected subset ($\Delta G_{i|\delta j}^{\text{stat}}$) is calculated:

$$\Delta\Delta G_{i,j}^{\text{stat}} = kT^* \sqrt{\sum_x \left(\ln \frac{P_{i|\delta j}^x}{P_{\text{MSA}|\delta j}^x} - \ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \tag{2}$$

This statistical coupling energy quantifies the degree to which the distribution of the amino acids at position i is modified by the perturbation in column j. In the second variant of SCA, the measure is reformulated as a weighted covariance:

$$\Delta\Delta G_{i,j}^{\text{stat}} \approx -\frac{1}{f_i^{(a_i)}} \cdot \frac{\partial D_j^{(b)}}{\partial f_j^{(b)}} \cdot C_{ij} \tag{3}$$

with $f_i^{(a_i)}$ being the frequency of amino acid type a_i at position i, $D_j^{(b)}$ the relative entropy that indicates how unlikely the observed frequency of amino acid b at position j was if b occurred randomly according to the underlying background frequency, and C_{ij} the covariance of the MSA columns i and j. For further details, see Ref. 3.

3 Mutual Information

An alternative to SCA is the Mutual information (MI). It is an information-theoretical measure based on Shannon's entropy[4]. It quantifies the dependence of two columns in an MSA by describing how much information about a residue in column i is obtained by revealing the type of the corresponding amino acid in column j.

The MI of an amino acid pair (i, j) is calculated by

$$\mathbf{M}_{ij} = \sum_{\sigma_i} \sum_{\sigma_j} p(\sigma_i, \sigma_j) \cdot \log_2 \left(\frac{p(\sigma_i, \sigma_j)}{p(\sigma_i) \cdot p(\sigma_j)} \right) \tag{4}$$

where σ is an alphabet of the 20 amino acids as well as characters for gaps ("-") and unknown types ("X"). $p(\sigma_i)$ and $p(\sigma_j)$ are the frequencies of amino acid types σ_i and σ_j in the respective columns.

To correctly handle gaps and unknown positions, one typically applies a null model normalization[5–7], which can be computed efficiently on GPUs[8].

4 Critical Assessment

4.1 Selection of Perturbations is Arbitrary

In the first formulation of SCA (Eq. 2) the choice of the type of the to-be-perturbed column j is rather arbitrary. For example, it is not clear which amino acid of column j to choose if
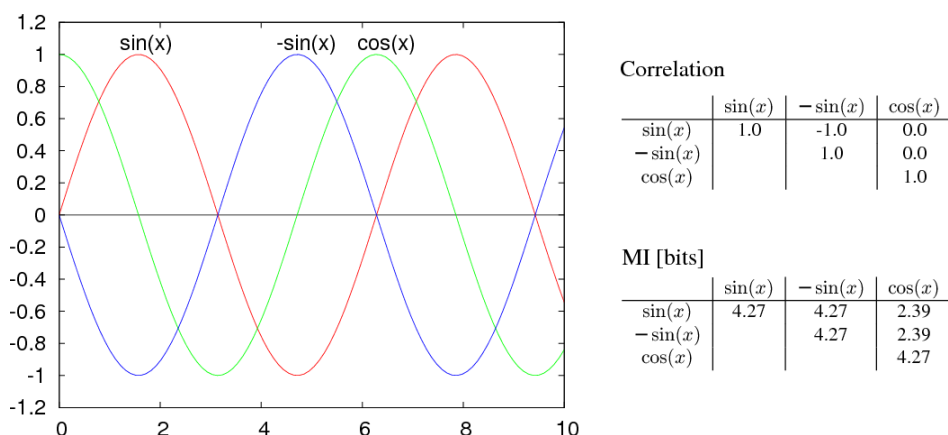
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2

sin(x)   -sin(x)   cos(x)

0   2   4   6   8   10

Correlation

	$\sin(x)$	$-\sin(x)$	$\cos(x)$
$\sin(x)$	1.0	-1.0	0.0
$-\sin(x)$		1.0	0.0
$\cos(x)$			1.0

MI [bits]

	$\sin(x)$	$-\sin(x)$	$\cos(x)$
$\sin(x)$	4.27	4.27	2.39
$-\sin(x)$		4.27	2.39
$\cos(x)$			4.27

Figure 1. Correlation and mutual information (MI) of the functions $\sin(x)$, $-\sin(x)$, and $\cos(x)$ shows that MI can reveal relationships which covariance cannot detect.

there is an empirical draw. At the same time, the reference (and therefore the evolutionary selection) might not be for a specific amino acid, but rather for a chemical character, such as hydrophobicity or charge. Such influence would be neglected, even discarded, in the original SCA. To cope with this, a more involved version of SCA was suggested, that is based on a weighted covariance (compare Eq. 3).

4.2 Covariance is not an Appropriate Measure for Coevolution

However, analyzing relationships by calculating the covariance as in Eq. 3 might lead to inconsistent results. As exemplified in Fig. 1 the $\sin(x)$, $-\sin(x)$ and $\cos(x)$ functions have a close relationship between them[a]. But when we take a look at their normalized covariance, the correlation, we can only see high absolute values between $\sin(x)$ and $-\sin(x)$ while $\cos(x)$ falsely seems to be independent of the other two functions. In contrast, the MI of Eq. 4 is able to reveal this connection. It is therefore a better measure for relationships of any two variables than approaches based on covariance.

4.3 Parametric Methods are Error-Prone

Approaches to examine coevolution can be classified into two basic groups: parametric and non-parametric methods. While non-parametric approaches work with the MSA as sole input, parametric procedures rely on models with additional assumptions. Due to our limited knowledge of the underlying processes of molecular evolution, these additional parameters are most likely inaccurate or even biased towards a certain subset. In SCA, the statistical coupling energy of an alignment column is computed from a subalignment including only those sequences that display the most conserved amino acid in the column

---

[a]rather intuitively: knowing the value of one function, one immediately can restrict the possible function values of any other of the three to just two possible values

233

under consideration. Thereby, information of the excluded sequences is lost and a selection bias is created.

Due to the reduction of the MSA to smaller subalignments, which is required for SCA, a larger number of sequences is necessary. For the same reason, SCA can only be performed for positions with a certain degree of conservation to ensure an appropriate size of the subalignment. This implies the parameterization problem from above, but also demands for much larger data sets, typically not available.

4.4 Conclusion

For these reasons we consider SCA as a historically important, but suboptimal approach to perform an unbiased coevolutionary analysis. It is designed as a perturbation analysis for examining carefully selected positions in an MSA. To conduct a general analysis, the method of MI in Eq. 4 should be applied instead.

Acknowledgments

References

1. F M Codoñer and M A Fares, *Why should we care about molecular coevolution?*, Evol Bioinform Online, **4**, 29–38, 2008.
2. A Bairoch and R Apweiler, *The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.*, Nucleic Acids Res, **28**, no. 1, 45–48, Jan. 2000.
3. S W Lockless and R Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families*, Science, **286**, no. 5438, 295–299, 1999.
4. C El Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J., **27**, no. 3, 379–423, 1948.
5. P Weil, F Hoffgaard, and K Hamacher, *Estimating sufficient statistics in coevolutionary analysis by mutual information*, Comput. Biol. Chem., **33**, no. 6, 440–444, 2009.
6. P Boba, P Weil, F Hoffgaard, and K Hamacher, "Intra- and inter-molecular coevolution: The case of HIV1 protease and reverse transcriptase", in: Biomedical Engineering Systems and Technologies, Ana Fred, Joaquim Filipe, and Hugo Gamboa, (Eds.), vol. 127 of *Springer Communications in Computer and Information Science*, pp. 356–366. Springer, Berlin Heidelberg, 2011.
7. K Hamacher, *Relating sequence evolution of HIV1-protease to its underlying molecular mechanics*, Gene, **422**, no. 1-2, 30–36, 2008.
8. M Waechter, K Hamacher, F Hoffgaard, S Widmer, and M Goesele, *Is Your Permutation Algorithm Unbiased for $n \neq 2^m$?*, in: 9th Int. Conf. on Parallel Processing & Appl. Mathematics (PPAM2011), Sept. 2011.

Towards the Description of In-cell Environments

Claudia Barrera-Patiño, Emiliano Ippoliti, Chuong H. H. Nguyen, Paul Strodel, Chao Zhang, Giulia Rosetti, and Paolo Carloni

Computational Biophysics Lab
German Research School for Simulation Sciences, 52425 Jülich, Germany,
Forschungszentrum Jülich, 52425 Jülich, Germany, and
RWTH Aachen University, 52056 Aachen, Germany
E-mail: p.carloni@grs-sim.de

The functionality of biological macromolecules is adapted to their inhomogeneous cellular environment. Meaningful theoretical descriptions of their properties therefore require detailed representations of their respective cellular compartment. In this contribution, we provide a brief survey of our studies on some important components present in cellular environments using various computational simulations approaches: (i) the dynamics of an excess proton close to a membrane surface, (ii) the permeation of small ions through a bacterial porin, (iii) fluorescent biosensors for the detection of *in vivo* phosphate and (iv) the dynamics of sensory rhodopsin II inside a membrane environment.

1 Introduction

Although biomolecules such as proteins and nucleic acids are the engines of the cell, our knowledge on how they behave in cells is limited. Cellular environments are extremely complex, with the macromolecule concentration as much as 400 mg/ml[1]. In addition, proteins function in a very heterogeneous medium, from cell membranes to the cytoplasm, which is a highly saline environment. Various components of cellular environment such as ions, membranes and crowding can considerably alter the reactivity of individual macromolecules, both qualitatively and quantitatively. Therefore, to identify the functional form of macromolecules in-cell conditions, these cellular effects must be taken properly into account.

In this paper, we briefly survey our computational studies on several cellular environmental factors (components), such as pH value (excess protons), ionic strengths (K^+, Na^+ ions), buffering agents (phosphate ions) and membranes (its integral proteins). We first present the study of a single excess proton migrating at water/hydrophobic interfaces. Next we talk about K^+ and Na^+ ion permeation through a bacterial porin. After that we report our progress on detecting in-cell inorganic phosphate by fluorescent biosensor molecules. In the last part we show how the integral protein - the sensory rhodopsin II - behaves in a membrane environment

2 Excess Proton at Water/Hydrophobic Interfaces

Water/hydrophobic interfaces are present ubiquitously in biological systems[2]. Recently, fluorescence experiments by our collaborator P. Pohl[3] along with multistate empirical valence bond (MS-EVB)[4] and polarizable force-field calculations[5] have lead to the proposal that excess protons may be localized close to such interfaces. The presence of a proton at
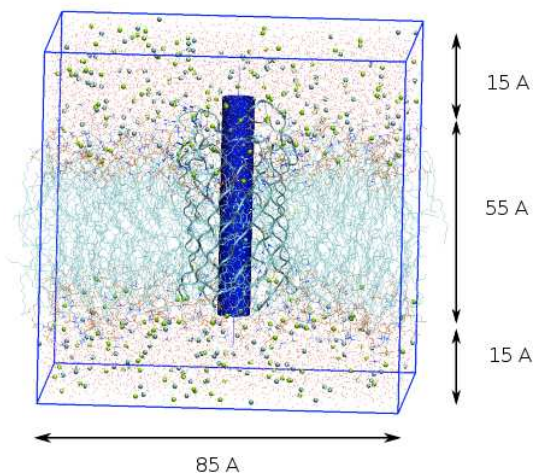
Figure 1. NanC periodic simulation system. A blue cylinder marks the ion passage path.

hydrophobic surfaces could play a pivotal role in a variety of fundamental interactions, including proton folding, protein-ligand (ligand= protein, DNA, RNA, drug) and lipid membrane/cytoplasm interactions[6]. We are applying Car-Parrinello (CP)[7] *ab initio* molecular dynamics (MD) simulations, which include electronic polarization and charge transfer effects to study the proton migration at such interfaces, applying the CPMD program[8].

Our system consists of 25 decane molecules, 302 water molecules and 1 excess proton (1707 atoms in total). The free energy (FE) profile of the excess proton in the direction perpendicular to the water/hydrophobic liquid surface has been provided by 70 ps *ab initio* metadynamics[9]. To the best of our knowledge, this is the largest first-principle based free energy calculation presented so far on a (bio) chemical system.

3   Ion Permeation through a Bacterial Porin

The permeation of small inorganic ions through cell membranes is a key biological element in physiological processes like nerve impulse propagation, cardiac and smooth muscle contraction[10]. Clearly, there are certain proteins, integrated in the membrane and spawning from the inner to the outer compartment, through which the ions pass. Important phenomena being investigated are (i) a given protein channel's selectivity for certain ion species and (ii) gating the opening and closing of the protein, induced by voltage changes or ligand binding. Trying to work on these topics[11] by means of atomistic simulations still is a challenge for many reasons, especially since only few 3D structures of protein channels have been resolved and system sizes for realistic simulations are huge, i.e highly demanding on computational resources.

Recently, the class of KdgM bacterial outer-membrane proteins was discovered[12], and for NanC, an acidic sugar-transporting porin, the structure was resolved[13]. Since NanC forms a compact β-barrel and allows for permeation of small ions like Na$^+$ or K$^+$, it is a suitable candidate for computational studies in terms of simulation sizes (cf. Fig. 1).

Also, by Bob Eisenberg's group, it was shown that NanC is highly efficient in conducting K^+ ions[14]. Thus, NanC is an ideal starting point to work on the mechanisms of Na^+ or K^+ permeation by means of MD simulations.

The permeation dynamics of the ions, here Na^+, K^+ and Cl^-, is determined by FE changes along the permeation coordinate (the channel axis). Starting from the crystal structure of NanC, we built an atomistic model, including the DOPC membrane and the water and ion environment (ca. 71000 atoms). We are using metadynamics[9] and umbrella sampling[15] to create FE profiles. Analysing the FE curves, in particular minima and barriers, yealds a deeper insight on (i) the protein atom groups strongly interacting with the traversing species, (ii) estimates about diffusion times and (iii) differences between Na^+ and K^+ permeation.

Using classical force field techniques in this context, however, does not allow us to calculate experimental observables related to selectivity, like current-voltage curves (CVCs) or ion fluxes (IFs). This is why we closely collaborate with Bob Eisenberg, who uses our structural information, e.g. data of type (i), as input parameters for other classes of calculations like Monte Carlo, Density Functional of Fluid-Poisson-Nernst-Planck theory[16] or energetic variational ones[17]. Since those simulations yield CVS and IFs directly, this combination of techniques allows us to verify the predictions from our FE calculations.

4 Detecting In-cell Phosphate by Fluorescent Biosensor

Fluorescence probes are routinely used to investigate the structural basis of biochemical processes *in vitro* and *in vivo*. Webb and co-workers developed efficient rhodamine-based fluorescent biosensors suitable for real-time measurements of inorganic phosphate (Pi)[18,19], exchanged in a multitude of cellular processes[20]. The sensor uses a protein very robust to denaturation as a scaffold, the phosphate binding protein (PBP) from *Escherichia coli*. Two rhodamines (6-IATR) have been covalently attached to mutation-generated cysteines (A17C, A197C) on the surface of the protein as fluorescent probes (Fig. 2). A17 and A197 have been selected to obtain a large change in the relative positions of the two fluorophores upon Pi binding[22,23]. As a result, the fluorescence emission of the rhodamines increases ∼18-fold upon binding Pi, in accordance with a reduction of the degree of stacking, which is known to quench fluorescence[19].

To further analyze the dependence of optical fluorophore properties on protein conformational changes, in particular the ability to promote stacking, mutant variants of rhodamine-PBP (A17C, A197C) were investigated. In particular, two additional mutations were introduced into the rhodamine-PBP (A17C, A197C) system: Y198A and L291A. These two amino acids are predicted from the calculated Pi-free structure to interact with the rhodamine closer to the protein surface, i.e. the rhodamine attached to C197. Structural models have been generated for rhodamine-PBP (A17C, A197C) including the mutations L198A and L291A both with and without Pi. These systems were equilibrated by classical MD methods using the NAMD program[24] in preparation of subsequent quantum mechanical treatment. The classical simulations show that Y198A interacts through its aromatic ring with rhodamine attached to C197 all over the classical trajectory, thus stabilizing the chromophores' binding. Based on our calculations, we therefore propose that Y198A rhodamine-PBP·Pi and rhoadmine-PBP will result in different structural adducts as compared to the parent system, which will affect the optical spectrum of the chromophores.
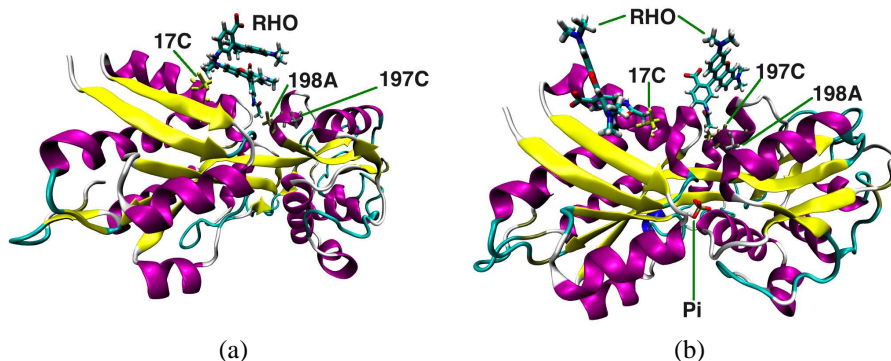
Figure 2. A17C, A197C RHO$_2$-PBP adducts with Y198A with (b,d) and without (a,c) Pi bound, constructed based on the X-ray structures of PBP in complex with Pi[18] and of T141D PBP[21], respectively.

For this goal we also introduced the L291A mutation. To determine the structural dependence of optical properties we are applying the CP/MM hybrid *ab initio*/classical MD approach[25, 26], combined with TDDFT and GW-BSE calculations, which has been already used by us for investigating absorption spectra of formaldehyde and indole[27–29] in aqueous solution. Such approach ensures the inclusion of the environmental effects in the optical properties.

5 Activation of Sensory Rhodopsin II in Explicit Membrane Environment

Throughout the microbial, animal, and plant world, organisms have evolved exquisitely sensitive photo detectors that monitor spatial and temporal variations in light intensity and color. Understanding the ability of living organisms to absorb visible photons is the essential issue of photobiology. Photoreceptors in *Halobacterium Salinarum*, a microbial organism, have been studied extensively because they provide simple model systems for the two fundamental functions of light sensing: active transport and phototaxis[30]. Active transport allows the organism to convert light energy into chemical energy while phototaxis modulates the swimming behavior of the bacterium in response to the light. These photoreceptors are all membrane proteins which share a common topology, namely the seven transmembrane helices connected to each other by protein loops. These seven helices form an interior binding pocket containing a retinal molecule, a form of vitamin A. Upon light excitation the retinal changes conformation and charge distribution which lead to activation of the receptor.

Sensory rhodopsin II in *Halobacterium Salinarum* (HsSRII) functions as a light-sensor that mediates the negative phototaxis from harmful blue-green light[31]. HsSRII is tightly bound to its cognate transducer protein HtrII, forming a complex inside the cell membrane. Upon light activation, HsSRII relays the signal to HtrII, which in turn initiates a cascade that regulates the cell's flagellar motor[32]. Thus, the light signal is converted to the motion

in *Halobacterium Salinarum*. Understanding how only few switches of the retinal can induce the structural change of the receptor is of the utmost interest for this photoreceptor.

Molecular dynamics (MD) simulations provide powerful tools to observe the dynamic fluctuation of protein structures with atomic resolution[33]. This approach allows us to study the conformational differences between distinct states of biomolecules. Therefore, we have been setting up two systems of HsSRII using the MD approach. One system is in the inactive conformation (dark state) and another one in an active conformation (M-state). The two structures were inserted into an explicit membrane environment and the MD simulations of those systems will be performed up to sub-microsecond timescales, using the program GROMACS[34]. Analysing the conformational differences between dark-state and M-state after finishing simulations will help us to provide new insight into the activation mechanism of HsSRII in membrane environment.

Acknowledgments

References

1. D. Homouz, M. Perham, A. Samiotakis, M. S. Cheung, and P. Wittung-Stafshede, *Crowded, cell-like environment induces shape changes in aspherical protein*, Proc. Natl. Acad. Sci. **105(33)**, 11754, 2008.
2. a) P. Ball, *Water as an Active Constituent in Cell Biology*, Chem. Rev. **108**, 74, 2008. b) D. Chandler, *Interfaces and the driving force of hydrophobic assembly*, Nature **437**, 29, 2005.
3. Y. N. Antonenko and P. Pohl, *Structural Proton Diffusion along Lipid Bilayers*, Biophys. J. **37**, 865, 2008.
4. S. Iuchi, H. Chen, F. Paesani and G. A. Voth, *Hydrated Excess Proton at Water-Hydrophobic Interfaces*, J. Phys. Chem. B **113**, 4017, 2009.
5. S. Vacha, D. Horinek, M. L. Berkowitz and P. Jungwirth, *Hydronium and hydroxide at the interface between water and hydrophobic media*, Phys. Chem. Chem. Phys **9**, 4736, 2007.
6. B. Alberts, A. Johnson, J. Lewis, M. Raff and K. Roberts, *Molecular Biology of the Cell New York: Garland Science*, (2008).
7. R. Car and M. Parrinello, *Unified apporach for molecular-dynamics and density-functional theory*, Phys. Rev. Lett. **55**, 2471, 1985.
8. IBM Corp 1990-2008, MPI für Festkörperforschung Stuttgart 1997-2001, *CPMD: 3.13.2 edition*, `http://www.cpmd.org/`, (2009).
9. A. Laio and M. Parrinello, *Escaping free-energy minima*, Proc. Natl. Acad. Sci. USA **99**, 12562, 2002.
10. B. Hille, *Ion Channels of Exitable Membranes*, (Sinauer Associates, 2001).
11. R. Mackinnon, *Potassium channels*, FEBS Lett. **555**, 62, 2003.

12. N. Blot, C. Berrier, N. Hugouvieux-Cotte-pattat, A. Ghazi, and G. Condemine, *The oligogalacturonate-specific porin kdgm of erwinia chrysanthemi belongs to a new porin family*, J. Biol. Chem. **277**, 7936, 2002.

13. C. Wirth, G. Condemine, C. Boiteux, S. Berneche, T. Schirmer, and C. M. Peneff, *NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KDGM porin family*, J. Mol. Biol. **394**, 718, 2009.

14. J. Giri, *Study of selectivity and permeation in voltage-gated Ion channels*, (PhD thesis, University of Chicago, Illinois, 2011).

15. G. M. Torrie, and J. P. Valleau, *Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling*, J. Comput. Phys. **23**, 187, 1977.

16. D. Gillespie, W. Nonner, and R. S. Eisenberg, *Coupling Poisson-Nernst-Planck and density functional theory to calculate ion flux*, J. Phys. Condens. Matter **14**, 12129, 2002.

17. P. Lin, and H. Zhang, *Simulations of singularity dynamics in liquid crystal flows: A C-0 finite element approach*, J. Comput. Phys. **215**, 348, 2002.

18. M. Brune, J. Hunter, J. Corrie, and M. Webb, *Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and Its application to actomyosin subfragment 1 ATPase*, Biochemistry **33(27)**, 8262, 1994.

19. M. P. Okoh, J. L. Hunter, J. E. T. Corrie, and M. R. Webb, *A biosensor for inorganic phosphate using a rhodamine-labeled phosphate binding protein*, Biochemistry **45(49)**, 14764, 2006.

20. M. R. Webb, *A fluorescent sensor to assay inorganic phosphate* in Kinetic analysis: a practical approach, pp 131-152. (Johnson, K. A., Ed., Oxford University Press, Oxford, UK, 2003).

21. N. H. Yao, P. S. Ledvina, A. Choudhary, and F. A. Quiocho, *Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor*, Biochemistry **35(7)**, 2079, 1996.

22. J. E. Selwyn, and J. I. Steinfel, *Aggregation of equilibriums of xanthene dyes*, J. Phys. Chem. **76(5)**, 762, 1972.

23. R. Chambers, T. Kajiwara, and D. Kearns, *Effect of dimer formation on the electronic absorption and emission spectra of ionic dyes. Rhodamines and other common dyes*, J. Phys. Chem. **78(4)**, 380, 1974.

24. J. C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot,R. D. Skeel,L. Kale, K. Schulten, *Scalable Molecular Dynamics with NAMD*, J. Comp. Chem. **26**, 1781, 2005.

25. A. Laio, J. VandeVondele, and U. Röthlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys. **116**, 6941, 2002.

26. M. Sulpizi, P. Carloni, J. Hutter, and U. Röthlisberger, *A hybrid TDDFT/MM investigation of the optical properties of aminocoumarins in water and acetonitrile solution*, Phys. Chem. Chem. Phys. **5(21)**, 4798, 2003.

27. A. M. Conte, E. Ippoliti, R. Del Sole, P. Carloni, and O. Pulci, *Many-Body Perturbation Theory extended to the Quantum Mechanics/Molecular Mechanics Approach: application to indole in water solution*, J. Chem. Theory Comput. **5(7)**, 1822, 2009.

28. P. Lupieri, E. Ippoliti, P. Alto, M. Garavelli, M. Mwalaba, and P. Carloni, *Spectroscopic properties of formaldehyde in aqueous solution: insights from Car - Parrinello and TDDFT/CASPT2 calculation*, J. Chem. Theory Comput. **6 (11)**, 3403, 2010.

29. A. M. Conte, E. Ippoliti, R. Del Sole, P. Carloni, O. Pulci, *Many-body meets QM/MM: Application to indole in water solution*, Phys. Status Solid B **247**, 1920, 2010.

30. J. L. Spudich and K. H. Jung, *Handbook of photosensory receptors*, (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, May 2005).

31. W. Zhang, A. Brooun, M. M. Mueller, and M. Alam, *The primary structures of the Archaeon Halobacterium salinarium blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein*, Proc. Natl. Acad. Sci. U.S.A. **93(16)**, 8230, 1996.

32. J. P. Klare, V. I. Gordeliy, J. Labahn, G. Büldt, H. J. Steinhoff, and M. Engelhard, *The archaeal sensory rhodopsin II/transducer complex: a model for transmembrane signal transfer*, FEBS Lett. **564(3)**, 219, 2004.

33. M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Mol. Biol. **9(9)**, 646, 2002.

34. C. Kutzner , D. van der Spoel, M. Fechner, E. Lindahl, U. W. Schmitt, B. L de Groot and H. Grubmüller, *Speeding up parallel GROMACS on high-latency networks*, J. Comput. Chem. **28(12)**, 2075, 2007.

Workshop Program

Wednesday 20.07.2011

07:30	Pick up at hotels in Jülich
08:00-08:45	Registration

Morning Session (Chair: Ulrich H.E. Hansmann)

08:45-08:55	Welcome by Prof. Dr. Dr. Thomas Lippert
08:55-09:00	Introductory remarks and introduction into session
09:00-09:20	Laudatio (Jeff Skolnick)
09:20-10:05	Harold Scheraga (Cornell University, Ithaca, NY, USA): Advances in Protein Biophysics
10:05-10:50	Adam Liwo (University of Gdansk, Gdansk, Poland): From Atomistic Simulations to Network Description of Biological Systems
10:50-11:15	Coffee break
11:15-12:00	Andrzej Kolinski (University of Warsaw, Warsaw, Poland): Coarse-grained Protein Modeling: Dynamics, Folding Pathways and Mechanical Unfolding
12:00-12:45	Panel discussion: Force fields Participants: H. Scheraga, Ch. Brooks, A. Kolinski, E. Lindahl, A. Liwo, U. Röthlisberger
12:45-14:00	Lunch

Afternoon Session (Chair: Olav Zimmermann)

14:00-14:05	Introduction to session
14:05-14:50	Richard Lavery (CNRS, Lyon, France): From Macromolecular Mechanics to Function
14:50-15:35	Kim Sneppen (Copenhagen University, Copenhagen, Denmark): Modeling Nucleosome Mediated Epigenetics
15:35-16:00	Coffee break
16:00-16:45	Michael Lässig (University of Cologne, Köln, Germany): From Biophysics to Fitness of a Molecular Pathway
16:45-17:15	Award talk: Marco Matthies (University of Hamburg, Hamburg, Germany): Continuous-space Sequence Optimisation for RNA Secondary Structures

17:15-18:00	Panel discussion: Molecular systems biology
	Participants: R. Lavery, M. Lässig, J. Skolnick, K. Sneppen
18:00	Welcome Reception and Poster Session
20:00	Bus to Hotels in Jülich

Thursday 21.07.2011

08:00   Pickup at Hotels in Jülich

Morning Session (Chair: Sandipan Mohanty)

08:30-08:35   Introduction to session

08:35-09:20   David Wales (Cambridge University, Cambridge, UK):
Exploring Energy Landscapes: From Molecules to Nanodevices

09:20-09:50   Award talk: Emal M. Alekozai (University of Heidelberg and Oak Ridge National Laboratory, Heidelberg, Germany):
Multilevel Enhanced Sampling of Cellulose-Cellulase Interaction

09:50-10:15   Coffee break

10:15-11:00   Erik Lindahl (KTH Royal Institute of Technology, Stockholm, Sweden):
Dynamics and Function of Voltage- and Ligand-gated Ion Channels

11:00-11:30   Award talk: Bettina Keller (Freie Universität Berlin, Berlin, Germany):
Markov State Models and Dynamical Fingerprints: Unraveling the Complexity of Molecular Kinetics

11:30-12:15   Panel discussion: Exploring/interpreting energy landscapes
Participants: H. Scheraga, H. Grubmüller, A. Kolinski, R. Lavery, D. Wales

12:15-12:30   Group photo

12:30-14:00   Lunch

Afternoon Session (Chair: Jan Meinke)

14:00-14:05   Introduction to session

14:05-14:50   Jeffrey Skolnick (Georgia Tech, Atlanta, USA):
Crowding and Hydrodynamic Interactions Likely Dominate *in vivo* Macromolecular Motion

14:50-15:35   Frank Jülicher (Max Planck Society, Dresden, Germany):
The Role of Pulling Forces in Cell Division

15:35-16:00   Coffee break

16:00-16:45	Amos Maritan (University of Padua, Padua, Italy):
	An Unified Perspective on Proteins: A Physics Approach
16:45-17:15	Award talk: Nicoletta Ceres (Université Lyon1 / CNRS, Lyon, France):
	PaLaCe: Toward a Transferable Coarse-grain Model for Protein Mechanics and Dynamics
17:15-18:00	Panel discussion: Cellular environments
	Participants: F. Jülicher, A. Liwo, A. Maritan, J. Skolnick
18:15	Bus to dinner at Castle Obbendorf (Hambach)
18:30	Dinner
22:00	Bus to Hotels in Jülich

Friday 22.07.2011

8:00	Pickup at Hotels in Jülich

Morning Session (Chair: Paolo Carloni)

08:30-08:35	Introduction to session
08:35-09:20	Helmut Grubmüller (Max Planck Society, Göttingen, Germany):
	Exploring Protein Dynamics Space: The Dynasome
09:20-10:05	Ursula Röthlisberger (Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland):
	Structural Stability of Prion Protein and Amyloid Beta Peptides and their Interaction with Metal Ions
10:05-10:35	Award talk: Yinghao Wu (Columbia University, New York, USA):
	Multiscale Simulations of Cadherin-mediated Cell Adhesion
10:35-11:00	Coffee break
11:00-11:45	Charles Brooks (University of Michigan, Ann Arbor, USA):
	Multi-scale Studies of Viral Capsid Dynamics and Mechanics
11:45-12:30	Panel discussion: Multi-scale modelling
	Participants: H. Scheraga, Ch. Brooks, H. Grubmüller, E. Lindahl, U. Röthlisberger
12:30-12:45	Concluding remarks
12:45-14:00	Lunch
14:00-15:00	Jülich Supercomputing Centre, Guided Tour
15:15	Bus to Cologne
	Social Event: Visit of a Traditional Brewhouse (drinks and food on one's own expense)
19:30	Bus to Jülich

List of Participants

- Adler, Thomas; University of Stuttgart; Wallerfangen; Germany

- Alekozai, Emal M.; University of Heidelberg and Oak Ridge National Laboratory; Heidelberg; Germany

- Barrera Patiño, Claudia Patricia; German Research School for Simulation Sciences; Jülich; Germany

- Bilsel, Mustafa; Ankara University; Ankara; Turkey

- Biswas, Mithun; University of Heidelberg; Heidelberg; Germany

- Blaszczyk, Maciej; University of Warsaw; Warsaw; Poland

- Brieg, Martin; Karlsruhe Institute of Technology; Eggenstein-Leopoldshafen; Germany

- Brooks, Charles; University of Michigan; Ann Arbor; USA

- Carloni, Paolo; German Research School for Simulation Sciences; Jülich; Germany

- Ceres, Nicoletta; Université Lyon1 / CNRS; Lyon; France

- Choy, Desmond; University of Cambridge; Cambridge; United Kingdom

- Cong, Xiaojing; International School for Advanced Studies; Trieste; Italy

- Corradi, Valentina; University of Calgary; Calgary; Canada

- Dibenedetto, Domenica; German Research School for Simulation Sciences; Jülich; Germany

- Dreyer, Jens; German Research School for Simulation Sciences; Jülich; Germany

- Eder, Christian; Forschungszentrum Jülich; Jülich; Germany

- Faber, Michael; Max Planck Institute of Colloids and Interfaces; Potsdam (OT Golm); Germany

- Falkner, Benjamin; Forschungszentrum Jülich; Jülich; Germany

- Ferro, Noel; University of Bonn; Bonn; Germany

- Finnerty, Justin; German Research School for Simulation Sciences; Jülich; Germany

- Giansanti, Andrea; Sapienza University of Rome; Rome; Italy

- Giorgetti, Alejandro; University of Verona; Verona; Italy

- Gront, Dominik; University of Warsaw; Warsaw; Poland

- Grubmüller, Helmut; Max Planck Society; Göttingen; Germany

- Grudinin, Sergei; INRIA; Saint Ismier; France

- Hamacher, Kay; TU Darmstadt; Darmstadt; Germany

- Hansmann, Ulrich; Michigan Technological University; Houghton; USA

- Heitzer, Christina; Philipps-Universität Marburg; Marburg; Germany

- Heuer, Andreas; WWU Münster; Münster; Germany

- Hoffmann, Falk; Forschungszentrum Jülich; Jülich; Germany

- Hung, Yu Fu; Forschungszentrum Jülich; Jülich; Germany

- Ippoliti, Emiliano; German Research School for Simulation Sciences; Jülich; Germany

- Jerabek, Hansjörg; Heidelberg Graduate School of Math. and Comp. Methods for the Sciences; Heidelberg; Germany

- Jülicher, Frank; Max Planck Society; Dresden; Germany

- Kamat, Sanjay; Forschungszentrum Jülich; Jülich; Germany

- Kamps, Martina; Forschungszentrum Jülich; Jülich; Germany

- Kar, Parimal; Max Planck Institute of Colloids and Interfaces; Potsdam; Germany

- Karlström, Gunnar; Lund University; Lund; Sweden

- Keller, Bettina; Freie Universität Berlin; Berlin; Germany

- Kirschner, Karl N.; Fraunhofer Institute; Sankt Augustin; Germany

- Klein, Doris Lenore; Heinrich-Heine Universität; Düsseldorf; Germany

- Klenin, Konstantin; Karlsruhe Institute of Technology; Eggenstein-Leopoldshafen; Germany

- Kmiecik, Sebastian; University of Warsaw; Warsaw; Poland

- Kolinski, Andrzej; University of Warsaw; Warsaw; Poland

- König, Bernd; Forschungszentrum Jülich; Jülich; Germany

- Kouza, Maksim; Michigan Technological University; Houghton, MI; USA

- Krammer, Eva-Maria; Université Libre de Bruxelles; Brussels; Belgium

- Kreyssig, Peter; University of Jena; Jena; Germany

- Krüger, Dennis M.; Heinrich-Heine-University; Düsseldorf; Germany

- Krüger, Jens; University of Paderborn; Paderborn; Germany

- Kulinski, Tadeusz; Polish Academy of Science; Poznan; Poland

- Lapidus, Dmitrijs; Latvian Institute of Organic Synthesis; Riga; Latvia

- Latek, Dorota; International Institute of Molecular and Cell Biology; Warsaw; Poland

- Lässig, Michael; University of Cologne; Köln; Germany

- Lavery, Richard; CNRS; Lyon; France

- Leguèbe, Michael; University Bordeaux 1 and ENSEIRB-MATMECA Bordeaux; Bordeaux; France

- Lindahl, Erik; KTH Royal Institute of Technology; Stockholm; Sweden

- Lippert, Thomas; Forschungszentrum Jülich; Jülich; Germany

- Liwo, Adam; University of Gdansk; Gdansk; Poland

- Margraf, Thomas; Uni Hamburg; Hamburg; Germany

- Maritan, Amos; University of Padua; Padua; Italy

- Matthies, Marco; University of Hamburg; Hamburg; Germany

- Meinke, Jan; Forschungszentrum Jülich; Jülich; Germany

- Minicozzi, Velia; University of Rome "Tor Vergata"; Rome; Italy

- Mohanty, Sandipan; Forschungszentrum Jülich; Jülich; Germany

- Musiani, Francesco; University of Bologna; Bologna; Italy

- Nadler, Walter; Forschungszentrum Jülich; Jülich; Germany

- Nguyen, Chuong H. H.; German Research School for Simulation Sciences; Jülich; Germany

- Nguyen, Trung Hai; German Research School for Simulation Sciences; Jülich; Germany

- Okamoto, Yuko; Nagoya University; Nagoya; Japan

- Olubiyi, Olujide; German Research School of Simulation Sciences; Jülich; Germany

- Osborne, Kenneth; Forschungszentrum Jülich; Jülich; Germany

- Pasi, Marco; BMSSI; Lyon; France

- Pisliakov, Andrei; RIKEN; Wako-shi; Japan

- Poojari, Chetan; Forschungszentrum Jülich; Jülich; Germany

- Quy, Vo Cam; German Research School for Simulation Sciences; Jülich; Germany

- Re, Suyong; RIKEN Advanced Science Institute; Saitama; Japan

- Roes, Claas; Forschungszentrum Jülich; Jülich; Germany

- Rossetti, Giulia; German Research School for Simulation Sciences; Jülich; Germany

- Röthlisberger, Ursula; Ecole Polytechnique Fédérale Lausanne; Lausanne; Switzerland

- Rückert, Chris; Forschungszentrum Jülich; Jülich; Germany

- Rudack, Till; Ruhr University Bochum; Bochum; Germany

- Rzepiela, Andrzej; Freiburg University; Freiburg; Germany

- Saini, Jagmohan; Heinrich-Heine-University; Düsseldorf; Germany

- Scheraga, Harold; Cornell University; Ithaca, NY; USA

- Schug, Alexander; Karlsruher Institut für Technologie; Eggenstein-Leopoldshafen; Germany

- Schweizer, Sabine; Technical University Munich; Freising; Germany

- Seo, Mikyung; University of Calgary; Calgary; Canada

- Shen, Jana; University of Oklahoma; Norman, Oklahoma; USA

- Shim, Vickie; University of Auckland; Auckland; New Zealand

- Sieradzan, Adam; University of Gdansk; Gdansk; Poland

- Singh, Gurpreet; University of Calgary; Calgary; Canada

- Singh, Sameer; Forschungszentrum Jülich; Jülich; Germany

- Skolnick, Jeffrey; Georgia Tech; Atlanta; USA

- Smiatek, Jens; WWU Münster; Münster; Germany

- Sneppen, Kim; Copenhagen University; Copenhagen; Denmark

- Strodel, Birgit; Forschungszentrum Jülich; Jülich; Germany

- Strodel, Paul; German Research School for Simulation Sciences; Jülich; Germany

- Sugita, Yuji; RIKEN; Saitama; Japan

- Tamamis, Phanourios; University of Cyprus; Nicosia; Cyprus

- Vancea, Ioan; Forschungszentrum Jülich; Jülich; Germany

- Wabik, Jacek; University of Warsaw; Warsaw; Poland

- Wales, David; Cambridge University; Cambridge; UK

- Wang, Zhe; Forschungszentrum Jülich; Jülich; Germany

- Weißgraeber, Stephanie; TU Darmstadt; Darmstadt; Germany

- Wieczorek, Grzegorz; Weizmann Institute of Science; Rehovot; Israel

- Wojtyczka, Andre; Forschungszentrum Jülich; Jülich; Germany

- Wu, Yinghao; Columbia University; New York; USA

- Zhang, Chao; German Research School for Simulation Sciences; Jülich; Germany

- Zimmermann, Olav; Forschungszentrum Jülich; Jülich; Germany

Schriften des Forschungszentrums Jülich
IAS Series

1. **Three-dimensional modelling of soil-plant interactions: Consistent coupling of soil and plant root systems**
   by T. Schröder (2009), VIII, 72 pages
   ISBN: 978-3-89336-576-0
   URN: urn:nbn:de:0001-00505

2. **Large-Scale Simulations of Error-Prone Quantum Computation Devices**
   by D. B. Trieu (2009), VI, 173 pages
   ISBN: 978-3-89336-601-9
   URN: urn:nbn:de:0001-00552

3. **NIC Symposium 2010**
   Proceedings, 24 – 25 February 2010 | Jülich, Germany
   edited by G. Münster, D. Wolf, M. Kremer (2010), V, 395 pages
   ISBN: 978-3-89336-606-4
   URN: urn:nbn:de:0001-2010020108

4. **Timestamp Synchronization of Concurrent Events**
   by D. Becker (2010), XVIII, 116 pages
   ISBN: 978-3-89336-625-5
   URN: urn:nbn:de:0001-2010051916

5. **UNICORE Summit 2010 – Proceedings**
   edited by A. Streit, M. Romberg, D. Mallmann (2010), iv, 123 pages
   ISBN: 978-3-89336-661-3
   URN: urn:nbn:de:0001-2010082304

6. **Fast Methods for Long Range Interactions in Complex Systems Lecture Notes**
   edited by P. Gibbon, T. Lippert, G. Sutmann (2011), ii, 167 pages
   ISBN: 978-3-89336-714-6
   URN: urn:nbn:de:0001-2011051907

7. **Generalized Algebraic Kernels and Multipole Expansions for massively parallel Vortex Particle Methods**
   by R. Speck (2011), iv, 125 pages
   ISBN: 978-3-89336-733-7
   URN: urn:nbn:de:0001-2011083003

8. **From Computational Biophysics to Systems Biology (CBSB11) – Celebrating Harold Scheraga's 90th Birthday**
   Proceedings, 20 – 22 July 2011 | Jülich, Germany
   edited by P. Carloni, U. H. E. Hansmann, T. Lippert, J. H. Meinke, S. Mohanty, W. Nadler, O. Zimmermann (2012), v, 251 pages
   ISBN: 978-3-89336-748-1
   URN: urn:nbn:de:0001-2011112819

The fifth workshop in the series "From Computational Biophysics to Systems Biology" (CBSB) took place from 20 to 22 July 2011 at Forschungszentrum Jülich. It was dedicated to Harold Scheraga who celebrated his 90th birthday that year. Dr. Scheraga pioneered the use of computers in chemistry and biology. His work inspired many of the research areas that are the topic of this meeting, ranging from biophysics to systems biology. In the spirit of Harold Scheraga's work, the workshop brought together researchers from physics, chemistry, biology, and computer science to acquaint each other with current trends in computational biophysics and systems biology, to explore avenues of cooperation, and to establish together a detailed understanding of cells at a molecular level.

These proceedings contain about 50 contributions from the CBSB11 workshop including a Laudation for Harold Scheraga by Jeff Skolnick and Harold Scheraga's biographic keynote lecture.

This publication was edited at the Jülich Supercomputing Centre (JSC) which is an integral part of the Institute for Advanced Simulation (IAS). The IAS combines the Jülich simulation sciences and the supercomputer facility in one organizational unit. It includes those parts of the scientific institutes at the Forschungszentrum Jülich which use simulation on supercomputers as their main research methodology.

JÜLICH

FORSCHUNGSZENTRUM