

Tertiary interactions in a model of the helix-coil transition

Cite as: J. Chem. Phys. **99**, 6865 (1993); <https://doi.org/10.1063/1.465831>

Submitted: 27 April 1993 . Accepted: 13 July 1993 . Published Online: 31 August 1998

Martin Ebeling, and Walter Nadler



View Online



Export Citation

Lock-in Amplifiers

Find out more today



Zurich
Instruments

Tertiary interactions in a model of the helix-coil transition

Martin Ebeling and Walter Nadler

*Institut für Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8,
D-72076 Tübingen, Germany*

(Received 27 April 1993; accepted 13 July 1993)

Into the classical Lifson–Roig model of the helix-coil transition in homo(poly)amino acids, we introduce a simple long-range interaction which models hydrophobic contacts between secondary structure elements, yet at the same time, retains its essentially 1D character. Although its properties cannot be determined analytically anymore, the model can be studied readily by Monte Carlo computer simulations. We show that our model assumptions lead to a markedly changed behavior in the transition from random coil to “folded” states. Remarkably, even in a homopolymer, the additional long-range interactions stabilize structures resembling secondary structure elements of globular proteins in number and average size. We present results from our simulations which demonstrate a glasslike transition behavior of the homopolymer model as well as an on-site construction mechanism for the formation of α helices.

I. INTRODUCTION

Since Anfinsen's¹ demonstration that, at least for some proteins, their complicated three-dimensional structure is completely encoded in their one-dimensional amino acid sequence, the question of how this coding is achieved and how a three-dimensional structure can be constructed from a knowledge of primary sequence alone has attracted ever growing interest.

This *protein folding problem* owes its intricacy to several factors, the foremost of which is the astronomically large number of conformations accessible to a protein chain made up of only a few dozen monomers, which renders an exhaustive search of conformations impossible.² Rather, during the very early stages of folding, the conformation space accessible to the chain has to be reduced substantially in size by means of interactions between monomers which are not too far apart from each other along the chain, and which, therefore, have a sufficiently large probability to encounter each other during a random search. The formation of secondary structures, notably of α helices, could serve that purpose.

However, such *local* interactions alone do not suffice to understand the protein folding problem, as has been shown by numerous attempts to predict secondary structure occurrence from primary sequence.^{3,4} The importance of *non-local* interactions between monomers which are spatially close to each other in the folded protein conformation, but far apart in the primary sequence is another factor which complicates the folding problem since it can lead to a very complicated free energy landscape. Various, structurally completely different, chain conformations may be energetically close to each other, but separated by high barriers of free energy giving rise to glasslike behavior with extremely long relaxation times for transitions between such minima.^{5,6}

In principle, the dynamical behavior of protein molecules can be addressed using the methods of molecular dynamics (MD) computer simulations.⁷ However, despite the enormous recent progress in this field, the sheer size of the systems under study still poses an unsurmountable

problem. Real-time simulations on even the fastest computers cover 10^2 to 10^3 ps at most, to be compared with real folding times in the range of seconds to minutes and longer.⁸

On the other hand, progress in the experimental study of the protein folding process has led to the description of identifiable intermediates formed within milliseconds during the folding process.^{9,10} Although these intermediates are rather ill defined structurally, they nevertheless tend to have a high content of secondary structure, at times even higher than that of the native state of the respective protein.¹⁰

For the study of the extremely relevant time scales in the nanosecond to millisecond range, there is, up to now, no better choice than to use simplified models. Such models necessarily represent compromises in that they should be easy to handle, yet realistic, at least, in their gross features. Here, the goal is to identify those features which allow to model proteinlike folding dynamics rather than to describe the folding of a particular protein. The most common approaches include the use of lattice models, use of stochastic instead of deterministic dynamics, and even a reduction of space dimension.^{11–16} In recent years, it has become clear that an analysis of such strongly simplified models can produce significant insights into the generic properties of the protein folding process.

Classical treatments in the history of simplified statistical-mechanical polypeptide models have been put forward by Zimm and Bragg,¹⁷ and by Lifson and Roig.¹⁸ In their original versions, these models describe the helix-coil transition in homo(poly)amino acids. Since only interactions between monomers separated by four positions along the chain are required to form α -helical structures, these models are essentially one-dimensional and can be solved analytically. Furthermore, they could be shown to reproduce the experimental data to a satisfactory degree.¹⁹ Therefore, it was felt that these models should be extended to the case of heteropolymers in order to be applicable to the dynamics of secondary structure formation in proteins. A vast amount of work has been dedicated to the formu-

lation of a heteropolymer model²⁰ and to the experimental determination of model parameters for the naturally occurring amino acids.²¹ However, as stated before, the description of a protein on the basis of local interactions alone is not adequate for the study of the folding process. Several attempts to incorporate an albeit simple form of tertiary interaction between α helices into the model have been undertaken.^{22,23} However, the requirement that the models should still be analyzable analytically, led to severe restrictions in the form of the interaction. Nevertheless, the ensuing analytical results were quite complex and could be analyzed numerically only.

In this paper, we present a model which retains the local interactions of the Lifson–Roig model, yet into which we introduce a simple form of tertiary interaction, similar in spirit to but more general than the ones suggested before. We have tried to maintain the essentially one-dimensional character of the earlier models. Due to the newly introduced interaction, the properties of our model cannot be determined analytically anymore. However, the model can be simulated easily already on small computers. Although the nonlocal interaction introduced here is among the simplest conceivable, and therefore, especially easy to handle computationally, it introduces a host of interesting new properties. In particular, it allows a stabilization of folded structure elements that are in some respects characteristic of folded globular proteins.

Since the formation and stabilization of secondary structure represents an important aspect of protein folding, various scenarios have been developed to describe the kinetics of these processes. In the *diffusion–collision* model,^{24,25} it is assumed that secondary structure elements form very early on in the folding process, undergo mutual diffusion, and are stabilized upon correct collision with each other. On the other hand, *on-site construction*^{26,27,15,16} models suggest the growth of secondary structure from

only a few initiation sites, during which the secondary structure elements stabilize each other already in the process of their formation.

To be able to distinguish between such different scenarios of the early folding stages, we have introduced into our model an additional feature which allows for a simplified mutual diffusion of secondary structure elements. Our aim, here, is to study the factors which favor or suppress diffusion–collision or on-site construction processes, respectively.

In the next section, we will introduce our model. In Sec. III, the ground state properties are analyzed in some detail. Simulation results for the folding/unfolding transition and for the low temperature dynamics of the model are represented in Sec. IV. Section V closes the paper with a brief summary and an outlook to heteropolymers.

II. DESCRIPTION OF THE MODEL

We will represent the conformation of an amino acid homopolymer of length L as a linear string in which each monomer i , $i=1, \dots, L$, is assigned a label $\sigma_i = h, c^+$, or c^0 , according to its local conformation. Following the classical treatment of the helix–coil transition in homo(poly)amino acids by Lifson and Roig,¹⁸ we denote by h any residue with dihedral angles within a range characteristic of residues in α helices. All nonhelical residues are assumed to be in “random coil” states and are labeled either c^+ or c^0 , the distinction between which will be clarified below. The existence of any secondary structure other than α helices is ignored here for simplicity. This is also a valid assumption for at least some globular proteins (of the “all- α ” type).

Choosing the all- c^+ chain as the reference state, the free energy of any chain conformation of length L with n helices of individual lengths l_i ($i=1, \dots, n$) can be cast into the form

$$F = -T \underbrace{\sum_{i=1}^L \Delta S(\sigma_i) + \Delta E_{HB} \sum_{i=2}^{L-1} H(\sigma_{i-1}, \sigma_i, \sigma_{i+1})}_{\text{local contribution}} + \underbrace{\sum_{i=1}^{n-1} C_{i,i+1} E_{i,i+1}^{\text{contact}}}_{\text{nonlocal contribution}} \quad (1)$$

The first, entropic, term is due to the restriction of conformation space of h or c^0 residues as compared to the conformation space of the reference c^+ residues. Denoting the conformation space volume of conformation σ_i by $V(\sigma_i)$, we obtain

$$\Delta S(\sigma_i) = k_B \ln \frac{V(\sigma_i)}{V(c^+)} \quad (2)$$

where k_B is Boltzmann's constant. Following the adaptation by Lifson and Roig¹⁸ of an experimental parameter first discussed by Zimm and Bragg,¹⁷ we set

$$V(h)/[V(c^+) + V(c^0)] = 0.0141,$$

which leads to

$$\Delta S(h)/k_B = -4.26 \ln[1 + \exp(\Delta S(c^0)/k_B)].$$

The second term in Eq. (1) represents the contribution from intrahelical hydrogen bonds that stabilize the helices. The establishment of an α helix requires the formation of at least one intrahelical hydrogen bond. Each hydrogen bond, in turn, requires three successive h residues along the chain.²⁸ In Eq. (1), this requirement is implemented by setting

$$H(\sigma_{i-1}, \sigma_i, \sigma_{i+1}) = \begin{cases} 1 & \text{if } \sigma_{i-1} = \sigma_i = \sigma_{i+1} = h; \\ 0 & \text{else.} \end{cases} \quad (3)$$

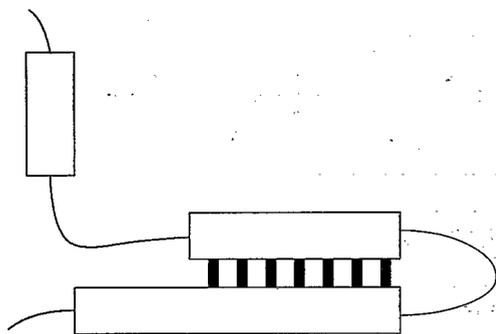


FIG. 1. Schematic representation of an amino acid homopolymer in which three distinct α -helical stretches have formed. The two larger helices are arranged in a hairpinlike structure which allows their amino acid side chains to interact with each other in the contact zone (indicated here by black bars between the helices).

Therefore, we obtain helices of length l ($l \geq 3$) with $l-2$ hydrogen bonds of energy $\Delta E_{HB} < 0$ each. Note that in random coil stretches, identified by the absence of intrahelical hydrogen bonds, some residues (or even pairs of residues) may, by chance, have dihedral angles characteristic of α helices and will then be labeled h , accordingly. However, these h residues will not contribute to the hydrogen bond term in Eq. (1) and will not be considered parts of helices.

Helices forming a tertiary contact between them stabilize each other mutually. In order to allow for such a contribution to the free energy, we have divided the nonhelical conformation space c of Lifson and Roig into two subspaces, introducing the random coil conformations c^0 and c^+ . We will assume that c^0 residues do not contribute to interhelical distances; in contrast, random coil residues c^+ , and isolated h residues (or pairs of them) in random coil stretches do contribute. Therefore, any two helices separated solely by c^0 residues are supposed to be in contact with each other.²⁹ If not otherwise stated, we assume the conformations c^+ and c^0 to have equal *a priori* probabilities, i.e., the respective conformation space volumes are equal, $V(c^+) = V(c^0)$. This yields $\Delta S(c^+) = \Delta S(c^0) = 0$ in Eqs. (1) and (2).

Tertiary interactions between helices can in a first approximation be considered to be proportional to the area of mutual contact.³⁰ In order to obtain a simple measure for the area of mutual contact, we assume for simplicity that two helices can adopt a hairpinlike structure in which they are supposed to be arranged "in register" as shown in Fig. 1.³¹ Thereby, the contact energy is determined by the length of the shorter helix ($k < 0$),

$$E_{ij}^{\text{contact}} = k \min(l_i, l_j). \quad (4)$$

The sum in Eq. (1) is over all helices, with l_i the length of the i th helix, and

$$C_{i,i+1} = \begin{cases} 1 & \text{if helices } i \text{ and } i+1 \text{ are in contact;} \\ 0 & \text{else.} \end{cases} \quad (5)$$

For $k=0$, the original Lifson-Roig model is recovered.

The contributions labeled as *local* in Eq. (1) are identical to those in the model of Lifson and Roig. Interactions within the chain which can depend on the conformations of stretches of arbitrary length—like the contact energy, Eq. (4)—can be called *nonlocal*. The introduction of such a nonlocal interaction renders it impractical at least, if not impossible, to evaluate the thermodynamic properties analytically.

However, as will be discussed below, the ground states of our model can easily be determined analytically. Thermodynamic and dynamic properties can be studied best by simple Monte Carlo simulations; in an elementary step, a monomer within the chain is chosen at random, and a random attempt to change this monomer's conformation is made which is accepted or rejected according to the Metropolis algorithm.³² By a Monte Carlo step (MC step) we will denote the number of elementary steps during which, on the average, each monomer within the chain is subjected once to an elementary step. For a polymer length of 100 residues, we have obtained an average simulation time of 0.1 s CPU time per Monte Carlo step on a μ VAX 3. Therefore, our model can easily be analyzed already on currently available personal computers.

The introduction of the conformation c^0 , which does not contribute to the distances between helices, gives rise to a very interesting feature in this model, i.e., the interconversion $c^+ \leftrightarrow c^0$ yields a one-dimensional model representation of the three-dimensional diffusional movement of individual helices against each other. Yet, it still allows the unambiguous representation of any chain conformation in a linear string of labels σ_j .

However simplified, the model presented here allows us to study how nonlocal interactions can influence the generic characteristics of the "folding process" of an amino acid polymer. Notably, since the model allows for diffusional movement of secondary structure elements, it can distinguish contributions of diffusion-collision processes to secondary and tertiary structure formation from on-site construction processes in the early phases of folding.

In the following, we will use a parameter b , defined by

$$b = -\frac{\Delta E_{HB}}{k_B T} + \ln \frac{V(h)}{V(c^+) + V(c^0)}, \quad (6)$$

as a dimensionless measure of temperature. Mark the *inverse* dependence of b on T , in analogy to $\beta = 1/k_B T$. This parameter closely resembles the quantity $\ln w$ introduced by Lifson and Roig; the transition point of the Lifson-Roig model ($k=0$) is near $b=0$.

In closing this section, we would like to shortly review approaches where interactions similar to Eq. (4) and Fig. 1 have been considered before. Closest to our model is probably Lauritzen and Zwanzig's description of two-dimensional polymer crystallization,³³ which they solved analytically for the thermodynamic limit (i.e., infinite chain length) case. Their essential result is a low-temperature phase exhibiting mutually interacting neighboring polymer strands of finite length. However, secondary structure formation, which we feel is essential for the application to protein folding, was disregarded in their ap-

proach. Poland and Scheraga,²² and later Skolnick,²³ considered systems with interhelical contacts where helical and coil secondary structure was taken into account. Both had to introduce various levels of constraints (only a given number of interacting helices, restrictions on helix and loop lengths and various different treatments of loops), so that the partition function could be described by either transfer matrix (for short chain lengths) or generating function (for $L \rightarrow \infty$) techniques. Moreover, their models differ from our model in that various parameters they have to use have no counterpart here. These considerations, and the unwieldy formalism they employed, do not make it appropriate, in our opinion, to compare their results to the ones we will present here.

III. FREE ENERGY MINIMA

A. Ground state conformations

In the low temperature range, the entropic term in Eq. (1) becomes small in comparison to ΔE_{HB} . In the limit of $k=0$, i.e., the Lifson-Roig case, our model shows a single, well-defined minimum of free energy which corresponds to a chain conformation with a single helix spanning the whole chain. However, by introducing the possibility of interhelical interactions, this situation is changed markedly. It turns out that now there are several different local free energy minima in the low temperature range which can be characterized by the number of helices in the chain. The corresponding chain conformations have some properties in common, irrespective of their numbers of helices. They are as follows.

- (i) All h residues are part of helices.
- (ii) The chain conformations contain only h and c^0 residues, since neither do c^+ residues take part in the formation of hydrogen bonds, nor do they allow the contact interaction between adjacent helices.
- (iii) As a consequence of (i) and (ii), all helices are in contact with their direct neighbors in these chain conformations; we refer to them as *compact conformations*.
- (iv) The loop between any two helices invariably consists of a single c^0 residue. Any longer loop would only shorten at least one of the adjacent helices, thereby reducing the number of hydrogen bonds and, possibly, the contact energy as well.

For any given number of helices, these properties already determine the local contributions to the free energy. In contrast, the contact energy contribution depends on the lengths of the individual helices. As a consequence, the conformation with the lowest free energy will be degenerate if the helical residues cannot be distributed evenly among the helices. This structural degeneracy is due to purely geometrical constraints arising from the condition that the total chain length L and the lengths of the n helices be integers, and will be referred to as $g(n;L)$. Table I shows $g(n;L)$ for chain lengths of about 100 and for helix numbers from 2 to 8. The results presented in Table I have been determined numerically. Note that, for a given number of helices, a very small change in chain length can result in a strong variation in the degree of degeneracy.

TABLE I. Degeneracy of local free energy minima.

| No. of helices | Chain length | | | | |
|----------------|--------------|----|-----|-----|-----|
| | 98 | 99 | 100 | 101 | 102 |
| 2 | 2 | 1 | 2 | 1 | 2 |
| 3 | 1 | 5 | 2 | 1 | 5 |
| 4 | 2 | 1 | 10 | 5 | 2 |
| 5 | 2 | 1 | 20 | 10 | 5 |
| 6 | 10 | 5 | 2 | 1 | 36 |
| 7 | 65 | 36 | 20 | 10 | 5 |
| 8 | 36 | 20 | 10 | 5 | 2 |

However, if a free energy minimum is degenerate, its free energy corresponds to that of a chain conformation for which the mean square deviation of the lengths of the individual helices from their average is a minimum.

The free energy at the local minima solely depends on the respective helix number n , and on the parameters L , k , and b (or, alternatively, T). Using the above properties, we obtain

$$\begin{aligned}
 F_{\min}(n;L,k,T) = & \Delta E_{HB}(L-3n+1) - (n-1)k_B T \Delta S(c^0) \\
 & - (L-n+1)T \Delta S(h) \\
 & + k \left\{ \left[\frac{L-n+1}{n} \right] (n-1) \right. \\
 & \left. + [(L-n+1) \bmod n] + \delta_{n,1} - 1 \right\} \\
 & - k_B T \ln g(n;L)
 \end{aligned} \quad (7)$$

for $(L-n+1)/n \geq 3$, which is the criterion that n helices can exist at all in a chain of length L . In contrast, the zero helix free energy minimum, i.e., the random coil state, comprises all the chain conformations made up solely from c^+ or c^0 residues. Due to its 2^L -fold degeneracy, its free energy is given simply by

$$F_{\min}(0;L,k,T) = -k_B T L \ln 2, \quad (8)$$

irrespective of the value of $\Delta S(c^0)$.

For any given temperature, we refer to the chain conformation corresponding to the global free energy minimum, i.e.,

$$F_{\min}(L,k,T) = \min_n (F_{\min}(n;L,k,T)), \quad (9)$$

as the *effective ground state* at this temperature.

B. Approximate phase diagram

From Eqs. (7) to (9), one can obtain the number of helices in the effective ground state for any values of the parameters L , b , and k . Using these results, one can construct an approximate phase diagram. This is done for a chain of length $L=100$ in Fig. 2. The respective helix numbers are indicated in the corresponding regions of the plot.

The phase diagram presented here is approximate in that it is based solely on the effective ground states. Figure

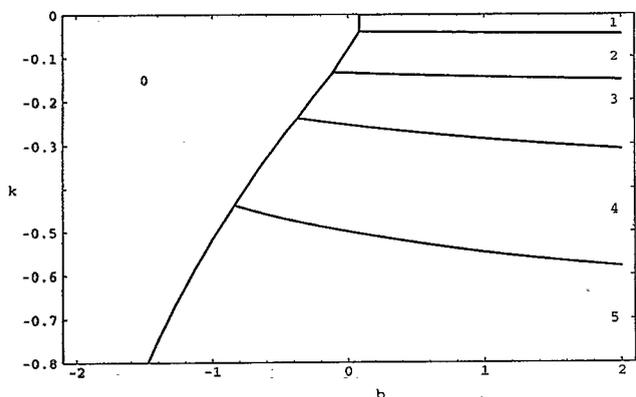


FIG. 2. Approximate phase diagram, based on Eqs. (7) through (9), $L=100$. Helix numbers in the effective ground state are indicated.

3 illustrates the problems involved. It shows the local free energy minima for chain conformations with two to seven helices for $L=100$, $k=-0.6$, and $b=0.5$. For this particular case, the number of helices in the effective ground state is five (compare Fig. 2). However, it is evident that the global minimum of free energy is *effectively* degenerate with respect to helix number, as the free energies of the conformations with four, five, and six helices, respectively, differ by less than 1%. Closer analysis shows that this effective energetic degeneracy holds over a wide range of b values.

In addition, for any given helix number, there are numerous energy levels (small dots in Fig. 3) which are energetically close to the respective local free energy minima. They correspond to chain conformations with small deviations from the optimum length distribution of the individual helices, with interhelical loops longer than one residue, or with coil residues at the chain ends leading to shortening of helices. These "excited states" are, as a rule, highly degenerate. Their degrees of degeneracy are extremely dif-

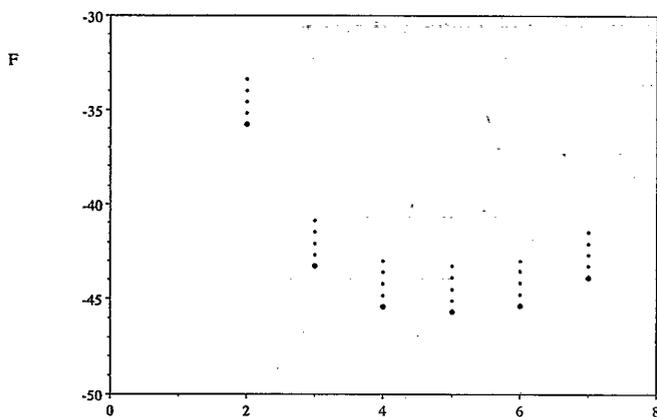


FIG. 3. (•) local free energy minima vs number of helices in the chain, $L=100$, $b=0.5$, $k=-0.6$; (○) some of the highly degenerate, "excited" free energy levels which arise from suboptimal length distributions among the helices.

icult to assess and, therefore, render it impracticable to evaluate the partition function of our model even approximately by an expansion around the free energy minima.

We stress, therefore, that the boundaries of the approximate phase diagram only correspond to those parameter values for which the local free energy minima [Eqs. (7) and (8)] for two different helix numbers become numerically identical. They are not to be viewed as separating pure thermodynamic phases. In thermal equilibrium, there will always be a mixture of states with different helix numbers present for finite values of b . However, at any point in the phase diagram, one can expect the helix number corresponding to the effective ground state to represent the maximum in the distribution of helix numbers in thermodynamic equilibrium.

Inspecting the approximate phase diagram, we find that for $0 > k > -0.04$, the boundary between random coil and helical states remains essentially at the value for the Lifson-Roig case, i.e., $b \approx 0.1$. For this range of k values, the chain conformation with a single helix spanning the whole chain is the only free energy minimum other than the random coil state. Below $k \approx -0.04$, the helix number for the effective ground state jumps to ever higher values. The boundaries mark the points where the energetic cost of disrupting a helix, which amounts to the breaking up of at least three intrahelical hydrogen bonds, is overcome by the accompanying gains in free energy. These are the gain in entropy for a newly introduced c^0 residue, and the gain in contact energy between the resulting daughter helices. For increasingly negative values of k , contacts between ever shorter helices, i.e., with ever smaller contact areas, are sufficient to compensate for the loss in hydrogen bonds. Therefore, effective ground states with ever shorter helices, i.e., with increasing helix numbers, are established.

Correspondingly, with increasingly negative values of k , the boundary between the random coil state and the respective helical states shifts to increasingly negative values of b . This is due to the strength of interhelical contacts which stabilize compact conformations at ever lower values of b , i.e., higher temperatures.

We have also studied the influence of the parameter $\Delta S(c^0)$, set to zero in Fig. 2, on the phase diagram. We find that over a wide range of values of $\Delta S(c^0)$, the boundary between the random coil subspace and the subspace of the compact conformations is only slightly altered. $\Delta S(c^0)$ has, however, a stronger influence on the boundaries between the various compact conformation subspaces. They are shifted towards smaller absolute values of k with increasing size of $\Delta S(c^0)$.

C. Implications for dynamics

Realistic dynamic steps in our model involve conformational changes of individual residues. This is true irrespective of the particular form of stochastic dynamics employed. Transitions between different free energy minima can be accomplished only via sequences of such elementary steps.

Figure 4(a) shows a detail from the free energy diagram of Fig. 3. The arrows in the diagram depict a typical

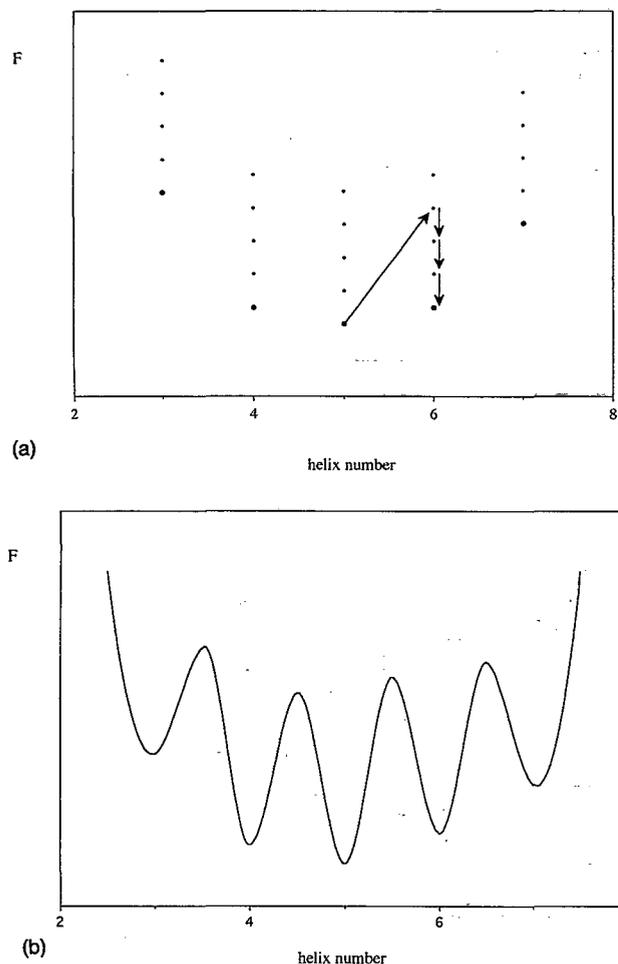


FIG. 4. (a) Detail of Fig. 3. The arrows indicate a possible route for the transition from the minimum with five helices to the one with six helices. (b) Schematic representation of free energy landscape to which transition pathways as in Fig. 4(a) between minima of free energy in effect amount.

way of moving from the local free energy minimum for five helices to the one for six helices. Even though many possible routes exist, the one shown in the figure serves to illustrate the fact that changes in helix number involve elementary steps that give rise to large free energy changes. For example, such steps break up helix contacts or helices. Therefore, free energy minima are separated by appreciable free energy barriers although the minima themselves may be energetically close to each other. This also holds for transitions to or from the random coil state, and although to a much lesser degree, for transitions between degenerate minima at a particular helix number. In effect, such constraints give rise to a rugged free energy surface, as shown schematically in Fig. 4(b).³⁴ Such a free energy landscape is typical for nonergodic, glasslike systems, a class to which proteins are well known to belong.³⁵ From these considerations, we can already conclude that our model will exhibit dynamic properties characteristic of such systems.

(i) The transition into the low-temperature phase depends strongly on the cooling velocity; in particular, *supercooling* occurs, and *hysteresis* upon reheating.

(ii) The model exhibits slow, anomalous relaxation in

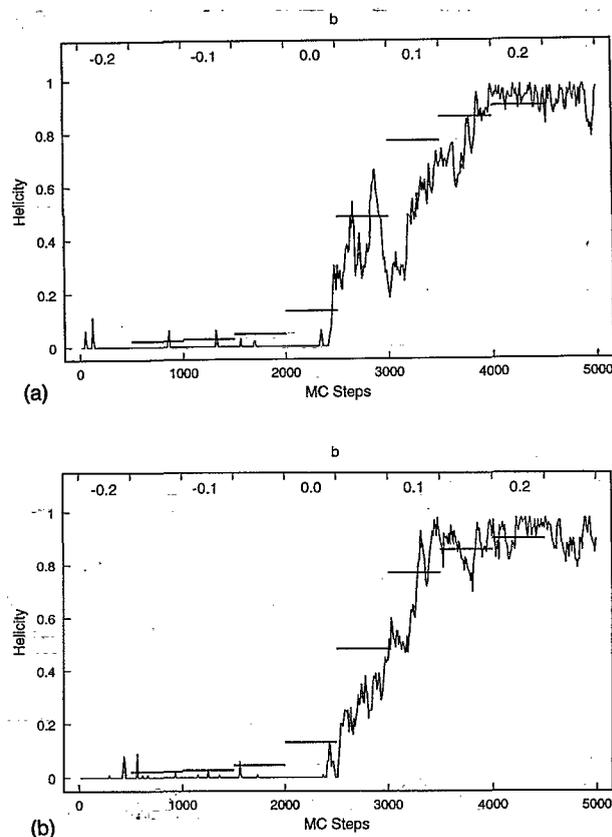


FIG. 5. (a), (b) Two individual folding trajectories for simulated annealing of a chain with $L=100$, $k=0$ (Lifson-Roig case). The value of b is raised by $\Delta b=0.05$ after 500 Monte Carlo steps each. For each particular value of b , 50 data points are shown, which represent averages over 10 MC steps each. Theoretical expectation values of helicity are indicated by vertical lines for every value of b .

the low-temperature phase, resulting in freezing of the system into different free energy minima.

These properties are investigated below in more detail.

IV. SIMULATION RESULTS

A. Coil-helix and helix-coil transitions

We study the temperature-induced helix-coil and coil-helix transitions of our model by simulated annealing. The cooling/heating schedule we use is simply an increase/decrease of b [Eq. (6)] by a constant value $\Delta b=0.05$ after N Monte Carlo steps each, which yields constant cooling, or heating, velocities $v = \pm \Delta b/N$.

Figures 5(a) and 5(b) show two cooling trajectories for the Lifson-Roig case ($k=0$). The states of the chains are characterized by their degree of intramolecular hydrogen bonding, i.e., by the average number of hydrogen bonds per residue.³⁶ The transition occurs near $b=0$ in both cases. Nevertheless, even at lower values of b , chain states with overall hydrogen bonding degree of up to 0.1 are sampled. Starting at $b=0$, the hydrogen bonding degree builds up gradually. From $b=0.1$, the state with a single helix spanning the chain clearly dominates all other chain states. Over the whole trajectory, the hydrogen bonding degree fluctuates strongly around the thermody-

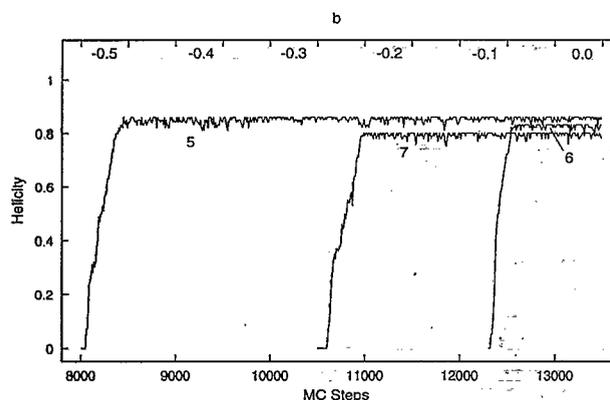


FIG. 6. Three individual folding trajectories for simulated annealing of a chain with $L=100$, $k=-0.6$. Starting from $b=-1.3$, the value of b is raised by $\Delta b=0.05$ after 500 Monte Carlo steps each. Representation as in Fig. 5. Trajectories are shown from where transitions start. Numbers in the plot give the respective helix numbers of the compact conformations to which the transitions lead.

dynamic expectation value as calculated from the partition function.¹⁸ For the higher b values (i.e., lower temperatures), it occasionally reaches its possible upper limit value.

Figure 6, in contrast, shows three cooling trajectories for the same cooling velocity as in Fig. 5, with the long-range interaction turned on. Several differences to Fig. 5 spring to attention. The transitions occur at various values of b , i.e., at $b \approx -0.5$, -0.25 , and -0.1 , respectively, in this case. These values are much lower than for the Lifson–Roig case, as to be expected from the phase diagram (Fig. 2). However, they still lie well within a region of the phase diagram where helical states are favorable over random coil states, i.e., *supercooling* occurs. The transitions are much steeper than those in Fig. 5. Each of them is well completed within 500 Monte Carlo steps. Throughout the trajectories, fluctuations are much less intense than in Fig. 5. The difference in final average hydrogen bonding degrees of the three trajectories shown is due to the fact that different numbers of helices, namely five, six, and seven, respectively, are established in the transitions. The intermittent c^0 residues between helices are not spanned by hydrogen bonds, which causes the overall hydrogen bonding degree to decrease to an extent depending on the number of helices established.³⁷ In Fig. 6, it can be seen that, although the degrees of hydrogen bonding do fluctuate somewhat after reaching a compact conformation, the number of helices is not changed for the three trajectories shown. The chains are frozen in different minima of free energy, and there is no efficient equilibration of chain conformations with similar free energies. Therefore, for the transitions shown here, we observe a distribution over states corresponding to various distinct free energy minima rather than an equilibrium mixture of states.

The altered transition behavior can be understood from the consideration that, even under conditions where a compact conformation represents the effective ground state, the initiation of secondary structure is *kinetically*

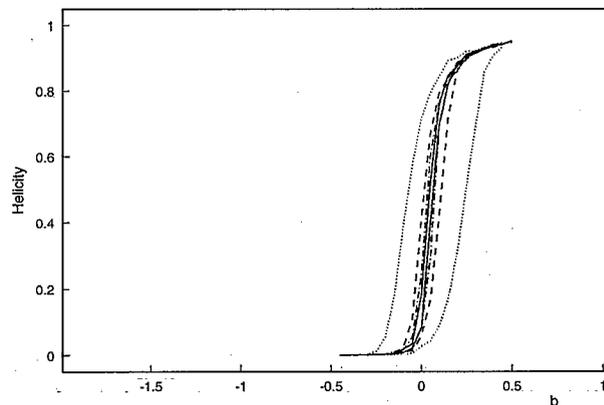


FIG. 7. Average cooling and heating curves for $L=100$, $k=0$, for various cooling or heating velocities. (dotted line) $v = \pm 5 \cdot 10^{-4}/\text{MC step}$, cooling ensemble size 100, heating ensemble size 100; (dashed line) $v = \pm 10^{-4}/\text{MC step}$, cooling ensemble size 100, heating ensemble size 100; (dashed-dotted line) $v = \pm 2 \cdot 10^{-5}/\text{MC step}$, cooling ensemble size 50, heating ensemble size 50; (straight line) $v = \pm 4 \cdot 10^{-6}/\text{MC step}$, cooling ensemble size 50, heating ensemble size 50.

hindered. Even when long-range interactions are included, the formation of any *first* helix is highly unlikely for $b < -0.05$, compare Fig. 5. Short helices can form occasionally due to thermal fluctuations, but are unstable, and, hence, short lived. Only rarely can two helices form which mutually stabilize each other by making an interhelical contact between them. In our simulations, we find that this is achieved by breaking up an elongated helix into two parts rather than by independent initiation of two helices. However, once such a nucleation of tertiary structure formation has occurred, the helices involved can grow quickly and break up into multiple helices, until a compact chain conformation has been reached (cf. Sec. IV C below). Therefore, the individual transitions are steeper than in Fig. 5, and pass through regions with intermediate hydrogen bonding degree in a very short time. These fast transitions do not sample the conformation space of the system efficiently. Rather, individual trajectories are attracted to a free energy minimum. Which particular minimum is selected depends on the random events leading to nucleation of secondary and tertiary structure formation during the transition. This also accounts for the fact that the transition points for individual chains are distributed randomly over a wide range of b values for nonzero cooling velocities. Only in the limit of infinitely slow cooling, one can expect to find a unique transition point.

Figures 7 and 8 show average cooling curves as well as the corresponding heating curves for ensembles of chains using different cooling/heating velocities. Figure 7 shows results for $k=0$, the Lifson–Roig case, while Fig. 8 gives the corresponding curves for $k=-0.6$.

For the Lifson–Roig case (Fig. 7), we find only little influence of cooling and heating velocities on the coil–helix and helix–coil transitions, respectively. For the slower velocities, the model behaves practically ergodic, i.e., the simulation runs using these velocities reproduce the thermodynamic expectation values for the hydrogen bonding

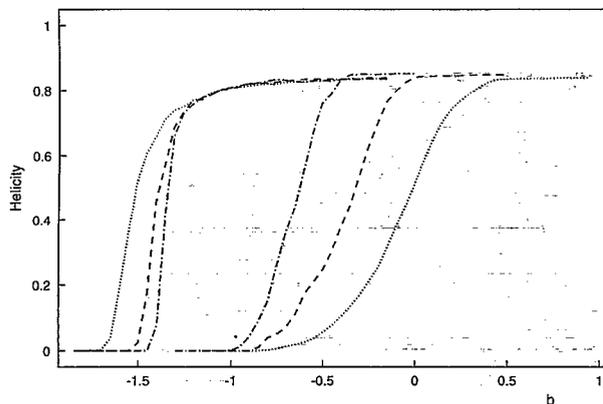


FIG. 8. Average cooling and heating curves for $L=100$, $k=-0.6$. (dotted line) $v = \pm 5 \cdot 10^{-4}/\text{MC}$ step, cooling ensemble size 250, heating ensemble size 70; (dashed line) $v = \pm 10^{-4}/\text{MC}$ step, cooling ensemble size 100, heating ensemble size 150; (dashed-dotted line) $v = \pm 2 \cdot 10^{-5}/\text{MC}$ step, cooling ensemble size 80, heating ensemble size 20.

degree. However, for the faster velocities, the cooling and heating curves markedly deviate from each other, which can be explained by the finite temperature jumps applied which cause the system to somewhat lag behind. For the same reason, the discrepancies of the cooling and heating curves, respectively, from the thermodynamic limit curve in Fig. 7 are practically symmetric for all velocities studied. The somewhat stronger deviations of the average cooling curves from the limit curve are due to late (i.e., $b > 0.05$) helix initiation events in some of the individual transitions in the ensemble.

In contrast, for the model with long-range interactions, the averaged transition curves show a markedly asymmetric behavior. The cooling curves (right side of Fig. 8) are strongly influenced by the cooling velocities; for slower cooling velocities, they are shifted towards lower values of b . For any single trajectory, the probability of helix initiations at a given value of b increases with increasing sojourn time at this value. This causes the distribution of transition points for individual chains to become narrower for slower velocities (i.e., longer sojourn times). Therefore, the average curves for the slower velocities are steeper in Fig. 8.

Once cooling has led to compact conformations, we observe *hysteresis* on subsequent heating. Note that the influence of the heating velocities on the helix-coil transition in Fig. 8 is much less drastic than that of the cooling velocities on the reverse transition. However, for ever slower velocities, the cooling and heating curves approach each other and would tend towards a single curve in the case of $v \rightarrow 0$. This common limit curve is impossible to assess from the simulations carried out. Its transition point should lie somewhere near $b \approx -1.2$ from the approximate phase diagram (Fig. 2). The relative discrepancies between the average heating curves in Fig. 8 (left side) are comparable to those in Fig. 7, whereas the average cooling curves differ much more from each other. Therefore, we assume that the average heating curve for the slowest heating velocity in Fig. 8 is already a good approximation to the

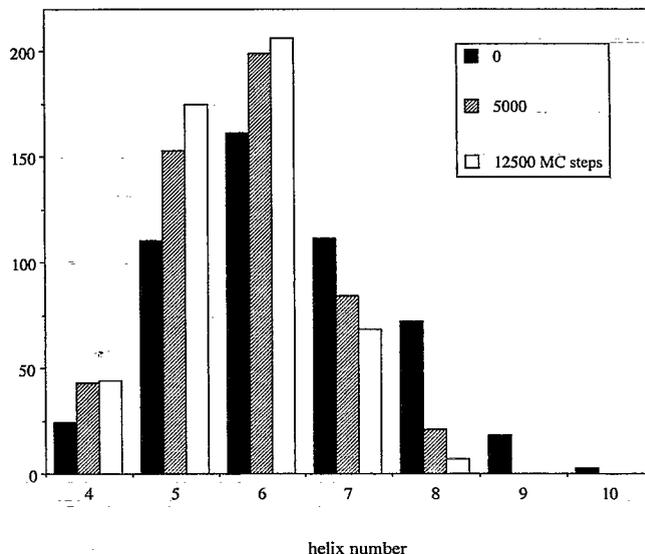


FIG. 9. Distributions of helix numbers for an ensemble of 500 individual chains ($L=100$, $k=-0.6$, $b=0.5$), for various simulation times (in MC steps after reaching a compact conformation).

thermodynamic limit curve for the coil-helix and helix-coil transition in our model.

Supercooling and hysteresis demonstrate that our model exhibits nonergodic behavior in the transition region.

B. Properties of the low temperature phase

In the preceding section, we saw that, for $k < 0$, the coil-helix transition under cooling leads to various compact conformations. These conformations correspond to local minima of the free energy which are energetically close to each other, but are separated by high energy barriers (compare Fig. 4). Due to these high barriers, a chain appears to be kinetically trapped in a local minimum, and changes in helix number are unlikely and, hence, rare events. For example, in Fig. 6, the trajectories with six and seven helices are stable with respect to helix number, although the effective ground state under the given conditions has five helices. Fluctuations in the degree of hydrogen bonding after completion of the transitions are mostly due to fluctuations of the individual helix lengths and at the chain ends.³⁸

For any particular trajectory, the minimum attained and the observables measured depend on the particular thermal fluctuations during the transition process. When simulating an ensemble of 500 chains ($L=100$, $k=-0.6$) at a constant $b=0.5$, starting from the all- c^+ state, we observed chain conformations with up to ten helices which lie energetically significantly higher than the states with four, five, or six helices.

Extending the simulations far beyond the point where a compact conformation had been reached for the first time, we observed extremely slow relaxation leading to a decrease in the average helix number, which was still in progress after 12 500 MC steps. Figure 9 shows distribu-

tions of the individual chains' helix numbers for various simulation times.

As a consequence, the equilibrium expectation values of observables in the folded phase, like number of helices, hydrogen bonding degree, and total free energy, cannot be obtained easily from simulation results.³⁹ In general, they are impossible to assess without evaluating the model's partition function. Only in the limit of $b \rightarrow \infty$ (i.e., $T \rightarrow 0$) can they be determined exactly; in this special case, the expectation values of the observables are identical to those of the global free energy minimum.

C. Folding kinetics

From what has been discussed so far, it is clear that the details of the coil-helix transition process deserve special attention. Analysis of the initiation of secondary structure and tertiary contact formation allows one to study the roles of diffusion-collision and on-site construction processes, respectively, and the factors which determine their relative importance. We believe that our simple computational model already has all the essentials to distinguish between various folding scenarios.

The particular form of our model, i.e., the representation of a chain conformation as a string of h , c^+ , and c^0 residues, allows the representation of a folding trajectory simply by a sequence of such strings. For an easier visual recognition, we represent h residues as *crosses*, c^0 as *dashes*, and c^+ as *blanks*. Figure 10 shows such a trajectory that starts from the all- c^+ state at $b=0$. Only a selection of chain states are shown, namely those resulting from changes in helix numbers or lengths. The numbers at the left give the elementary simulation step number.

More than half of the total simulation time is spent with formation and disruption of a total of 11 short (maximum length six residues), unstable helices which are all destroyed after an average lifetime of 138 elementary steps (not shown in Fig. 10). Since helix initiation is a rare event, the coexistence of two or more helices in a chain of length $L=100$ is very unlikely at this stage. The shown trajectory starts with the formation of a helix (of minimum length three residues) after 10 241 elementary steps. Inspection of the trajectory shows that, even where there are several helices present, their average lifetime is too short to allow for a diffusion-collision mechanism (elementary steps 11 314 to 11 359). After about 11 500 elementary steps, the first tertiary contact is formed by the breaking of a helix of ten residues. From there on, the helices grow and, once they have attained a certain length, break up again, thereby forming a steadily growing compact cluster. The average length of a breaking helix is about 22 residues. Since the helices do not usually split into fragments of equal length, which would be energetically desirable, there is a tendency to fine tune the helix lengths by a consecutive extending and shortening of the "loop" of c^0 residues between the helices, notably without ever breaking the contact between them. It is this mechanism which allows any compact chain conformation to reach a local free energy minimum without changing its helix number. In elementary step 17 183, a second nucleation site is formed by

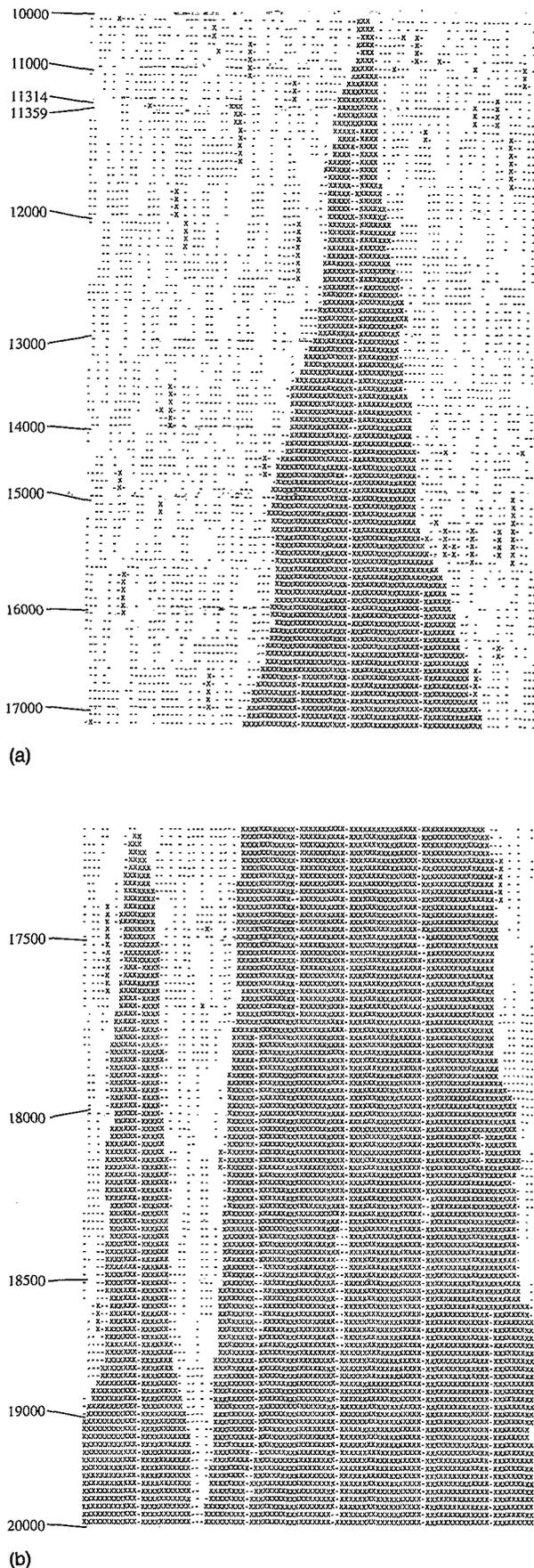


FIG. 10. (a), (b) Folding trajectory ($L=100$, $k=-0.6$, $b=0.5$); h residues are represented as *crosses*, c^0 residues as *dashes*, and c^+ residues as *blanks*.

initiation of a short helix independent of the compact, "folded" part of the chain. Note how the two nucleation centers are stabilized by interhelical contacts, i.e., on-site construction processes, long before they finally coalesce. Even then, they merge by growing towards each other rather than by a diffusion-collision process. It is when only three random coil residues are left between them that they first establish an interhelical contact. A compact state with six helices is reached shortly thereafter (not shown).

Since, in a homopolymer, initiation of helices and of tertiary contacts can occur at any location in the chain with equal probability, there is no way to establish a unique folding pathway. However, it will be very interesting to check heteropolymers for specific folding pathways, and to elucidate the events influencing the decision which compact conformations will be reached in the end. If on-site construction processes dominate the early stage of protein folding, then local preferences in a heteropolymer will be important for determining folding pathways. Recent experimental work, indeed, finds evidence for a prepartitioning of polypeptide chain conformation space by local propensities.^{40,41}

Most of the results presented in this paper have been obtained for a special set of parameters ($L=100$, $\Delta S(c^0)=0$, $k=-0.6$) using the Monte Carlo Metropolis dynamics presented in Sec. II. We would like to point out that, for these parameters at least, an on-site construction mechanism is clearly followed in the coil-helix transition. A more detailed investigation of the folding kinetics under various conditions is under way.

V. SUMMARY AND OUTLOOK

Helix-coil transitions in homo(poly)amino acids can be adequately described by the model of Lifson and Roig.¹⁸ However, one of the shortcomings of this model when trying to extend it to the case of heteropolymers is that it does not allow for the formation of tertiary structure. In the low temperature limit, a single helix spanning the whole chain is invariably established. Work by Poland and Scheraga,²² and by Skolnick,²³ has shown that tertiary, and, hence, nonlocal (in the sense used here), interactions can only be incorporated into the model under strongly limiting conditions if one insists that it still be solvable analytically.

In this paper, we show that a simple extension of the Lifson-Roig model introducing diffusional movement of secondary structure elements and tertiary interactions between them, can be studied already on a personal computer using Monte Carlo simulations. We believe that the resulting model already incorporates essential aspects required to analyze *generic* folding properties of proteinlike chain molecules. At the same time, it is easily handable and does not require time-consuming computer runs.

Most notably, even in a homopolymer, chain conformations are generated which resemble globular proteins with respect to secondary structure content, number of helices, and average helix length. The system shows non-ergodic, glasslike behavior in the low temperature, folded phase. Metastable chain conformations can be trapped ki-

netically, and relaxation to the effective ground state is extremely slow.

During the transition from random coil to helical structures, we do not observe the formation of multiple nucleation centers which upon diffusion collide and stabilize each other, as postulated by diffusion-collision models. Rather, we find rapid growth of secondary structure elements from only very few nucleation centers, a process which has been appropriately termed "on-site construction".^{26,27,16} The rapid formation of secondary structure with subsequent, slower structural rearrangements reminds of the "molten globule" intermediate in the folding of many globular proteins.⁹ Not surprisingly for a homopolymer, our model shows neither a genuine "folding pathway" nor a native structure.

The model presented here can be extended readily to the case of heteropolymers by employing sets of parameters $\Delta E_{HB}(A)$, $\Delta S(\sigma_i, A)$, and $k(A)$, where A stands for any type of monomer. It is conceivable that, at least for some sequences of monomers, the local or nonlocal specificities will remove the degeneracy of free energy minima observed for the homopolymer, thus leading to a unique "folded" structure. Local and nonlocal specificities may also have an influence on possible folding pathways and the balance of diffusion-collision vs on-site construction scenarios. We are currently investigating this possibility.

ACKNOWLEDGMENT

M. E. gratefully acknowledges support by a stipend from the Studienstiftung des Deutschen Volkes.

¹ C. B. Anfinsen, *Science* **181**, 223 (1973).

² C. Levinthal, in *Mössbauer Spectroscopy in Biological Systems*, edited by Debrunner *et al.* (University of Illinois, Urbana, 1969); however, for a somewhat different point of view, see R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. USA* **89**, 20 (1992).

³ *Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. Fasman (Plenum, New York, 1989).

⁴ N. Quian and T. J. Sejnowski, *J. Mol. Biol.* **202**, 865 (1988).

⁵ J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84**, 7524 (1987).

⁶ E. I. Shakhnovich and A. M. Gutin, *Biophys. Chem.* **34**, 187 (1989).

⁷ J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University, New York, 1989).

⁸ See Ref. 7 and references therein.

⁹ H. Christensen and R. H. Pain, *Eur. Biophys. J.* **19**, 221 (1991).

¹⁰ S. E. Radford, C. M. Dobson, and P. A. Evans, *Nature* **358**, 302 (1992).

¹¹ H. Taketomi, Y. Ueda, and N. Gö, *Int. J. Pept. Protein Res.* **7**, 445 (1975).

¹² H. S. Chan and K. A. Dill, *Proc. Natl. Acad. Sci. USA* **87**, 6388 (1990).

¹³ J. Skolnick and A. Kolinski, *Science* **250**, 1121 (1990).

¹⁴ E. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, *Phys. Rev. Lett.* **67**, 1665 (1991).

¹⁵ A. Rey and J. Skolnick, *Chem. Phys.* **158**, 199 (1991).

¹⁶ K. A. Dill, K. M. Fiebig, and H. S. Chan, *Proc. Natl. Acad. Sci. USA* **90**, 1942 (1993).

¹⁷ B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 526 (1959).

¹⁸ S. Lifson and A. Roig, *J. Chem. Phys.* **34**, 1963 (1961).

¹⁹ P. Doty and J. T. Yang, *J. Am. Chem. Soc.* **78**, 498 (1956).

²⁰ N. Gö, P. N. Lewis, M. Gö, and H. A. Scheraga, *Macromolecules* **4**, 692 (1971).

²¹ The theoretical foundation for the experimental determination of helix parameters was first laid by P. H. von Dreele, D. Poland, and H. A. Scheraga, *Macromolecules* **4**, 396 (1971). By 1990, parameters for all

- naturally occurring amino acids had been published, see J. Wojcik, K.-H. Altmann, and H. A. Scheraga, *Biopolymers* **30**, 121 (1990). However, recent work done by other groups has cast some doubt on the validity of the methods employed and on the values obtained: P. C. Lyu, M. I. Liff, L. A. Marky, and N. R. Kallenbach, *Science* **250**, 669 (1990); K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990); S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, and R. L. Baldwin, *Nature* **344**, 268 (1990).
- ²²D. C. Poland and H. A. Scheraga, *Biopolymers* **3**, 275 (1965); **3**, 283 (1965); **3**, 305 (1965); **3**, 315 (1965); **3**, 335 (1965); **3**, 357 (1965).
- ²³J. Skolnick, *Macromolecules* **16**, 1069 (1983); **16**, 1763 (1983); **17**, 645 (1984); **18**, 1073 (1985); **19**, 1153 (1986).
- ²⁴M. Karplus and D. L. Weaver, *Nature* **260**, 404 (1976).
- ²⁵D. Bashford, F. E. Cohen, M. Karplus, I. D. Kuntz, and D. L. Weaver, *Proteins* **4**, 211 (1988).
- ²⁶R. R. Matheson and H. A. Scheraga, *Macromolecules* **11**, 819 (1978).
- ²⁷A. Sikorski and J. Skolnick, *J. Mol. Biol.* **212**, 819 (1990).
- ²⁸Three successive h residues at positions $i-1$, i , and $i+1$ are spanned by a hydrogen bond, which connects the residues at positions $i-2$ to $i+2$, respectively. These latter two residues, however, are not restricted to α -helical dihedral angles; they are therefore not counted as members of the resulting helix. Thus, three h residues in a row are necessary for establishing a first intrahelical hydrogen bond. Any further h residue then adds another hydrogen bond to the helix.
- ²⁹We note that any two helices which are not direct neighbors along the chain are always separated by the residues of at least one intermittent helix, and thus, can never be in contact with each other. This is clearly an over simplification. However, such a "neighborhood correlation" is to be expected, at least, for the early phases of the folding process [P. J. Flory, *J. Am. Chem. Soc.* **78**, 5222 (1956)].
- ³⁰C. Chothia, *Nature* **248**, 338 (1974).
- ³¹Closely similar model assumptions for interhelical interactions have been introduced by Poland and Scheraga (Ref. 22), and by Skolnick (Ref. 23).
- ³²N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ³³R. Zwanzig and J. I. Lauritzen, *J. Chem. Phys.* **48**, 3351 (1968); J. I. Lauritzen and R. Zwanzig, *ibid.* **52**, 3740 (1970).
- ³⁴Of course, only integer helix numbers are possible.
- ³⁵D. L. Stein, *Proc. Natl. Acad. Sci. USA* **82**, 3670 (1985).
- ³⁶Following Lifson and Roig (Ref. 18), we assume two additional partners for hydrogen bonding to be placed at both chain ends, which allow $L-2$ hydrogen bonds to be formed in a chain of length L ; these additional residues are not taken into account further.
- ³⁷This is also the reason why the transitions in Figs. 5 and 7 yield higher final values of average helicity than those in Figs. 6 and 8. The latter ones lead to compact chain conformations with more than one helix.
- ³⁸Such fluctuations allow to find the conformation with the lowest free energy inside a free energy well corresponding to a particular helix number.
- ³⁹We are currently investigating the possibility to overcome such problems by using new simulation techniques developed for this purpose, see, e.g., E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992), and references therein.
- ⁴⁰H. J. Dyson, G. Merutka, J. P. Waltho, R. A. Lerner, and P. E. Wright, *J. Mol. Biol.* **226**, 795 (1992).
- ⁴¹H. J. Dyson, J. R. Sayre, G. Merutka, H.-C. Shin, R. A. Lerner, and P. E. Wright, *J. Mol. Biol.* **226**, 819 (1992).