LU-TP 97-04
March 26, 1997

On the Kinetic Behavior and Folding Properties of an Off-Lattice Heteropolymer Model

Ola Sommelius[1]

Complex Systems Group, Department of Theoretical Physics
University of Lund, Sölvegatan 14A, S-223 62 Lund, Sweden

Submitted to *Europhysics Letters*

Abstract:
The kinetic behavior of a three-dimensional off-lattice heteropolymer model is studied in terms of the time dependence of the average mean-square displacement between configurations. It is found that at short time-scales similar behavior is obtained even for sequences with very different thermodynamic properties. Furthermore, the degree of cooperativity in the folding process is examined by studying the residual number of degrees of freedom, obtained from an eigenvalue analysis of the correlation matrix, contributing to the structural fluctuations. In the compact state, a gradual decrease in this effective number of degrees of freedom take place as the temperature is lowered. This can be interpreted as an increasing asymmetry of the energy landscape.

[1] ola@thep.lu.se

1 Introduction

In order to understand the characteristics of the energy landscape and to quantify the requirements needed for a protein to have a thermodynamical stable yet kinetically accessible native state, the introduction of simple models [1] is necessary. In these models the small length and time scales are effectively averaged out and the thermodynamics of the resulting coarse grained chain is described by an effective potential. The basic assumption for this is that a low-resolution description still captures features essential for describing folding properties.

Models which have been much studied are the lattice models where the protein is represented as chain of beads on a cubic lattice (see references in [1]). This approach has indeed proven to be very useful and one is able to describe several non-trivial aspects in the folding process. However it is important to study alternative off-lattice models both in order to understand the limitations of the lattice models and in their own right. In this work a 3D off-lattice model suggested in [2] is used. The model contains two types of amino acids, hydrophobic and hydrophilic, and the formation of a hydrophobic core is induced by a sequence dependent Lennard-Jones potential. In [2] the studies were focused on the thermodynamical behavior of this model.

For a sequence to be a good folder it must satisfy both thermodynamic and kinetic requirements. That is the ground state must be stable against thermodynamical fluctuations, which happens at low temperatures, and yet be kinetically accessible, which happens at higher temperatures. Therefore only sequences for which the thermodynamic stability persists at high enough temperatures can be classified as good folders. To understand the behavior of the structural fluctuations in this context the overall shape diffusion at short times is studied as well as the collective behavior of the conformational correlations in the chain.

This letter is organized as follows. Section two defines the model and the observables. Section three and four contain the results and a brief summary respectively.

2 The model

The model contains two kinds of residues, A and B, which behave as hydrophobic (σ_i ='A') and hydrophilic (σ_i ='B') residues, respectively. These monomers are joined by rigid bonds $\hat{b}_i$ to form a linear heteropolymer chain living in three dimensions. The shape of the chain is thus specified by the $N-1$ bond vectors $\hat{b}_i$. The energy function is given by

$$E(\hat{b};\sigma) = -\kappa_1 \sum_{i=1}^{N-2} \hat{b}_i \cdot \hat{b}_{i+1} - \kappa_2 \sum_{i=1}^{N-3} \hat{b}_i \cdot \hat{b}_{i+2} + \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} E_{\text{LJ}}(r_{ij};\sigma_i,\sigma_j) \tag{1}$$

where $r_{ij} = r_{ij}(\hat{b}_i,\ldots,\hat{b}_{j-1})$ denotes the distance between sites i and j of the chain, and $\sigma_1,\ldots,\sigma_N$ is a binary string that specifies the primary sequence. The *species-dependent* global interactions are given by the Lennard-Jones potential,

$$E_{\text{LJ}}(r_{ij};\sigma_i,\sigma_j) = 4\epsilon(\sigma_i,\sigma_j)\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right) . \tag{2}$$

The depth of the minimum of this potential, $\epsilon(\sigma_i,\sigma_j)$, is chosen to favor the formation of a core of A residues, *i.e.* $\epsilon(A,A)=1$ and $\epsilon(A,B)=\epsilon(B,B)=1/2$. The two parameters of the energy function, κ_1 and κ_2, determine the strength of *species-independent* local interactions. The thermodynamic behavior of this model has been studied in Ref. [2]. It was found that the values $(\kappa_1, \kappa_2) = (-1, 0.5)$ give rise to local correlations qualitatively similar to those found in functional proteins. This choice favors anti-parallel nearest neighbour bonds and parallel next nearest neighbour bonds and will be used throughout this work. Furthermore it was found that, as the temperature is lowered, a gradual compaction occur. In the compact state this is then followed by a sequence dependent folding transition. For this model only a small fraction of random sequences have good folding properties. In 2D [3], where a more extensive investigation was done, only around 10 percent satisfied the folding criteria. This model to some extent resembles the IMP model [4, 5] which is a Gaussian chain augmented with a Lennard-Jones potential with an additional quenched disorder term $\sqrt{\epsilon}\,\eta_{ij}/r_{ij}^6$ representing a species-dependent interaction. The η_{ij}'s are stochastic variables having zero mean and unit variance, while ϵ is a measure of the strength of the quenched disorder.

In order to study the kinetics on a rugged energy landscape some kind of distance measure between conformations has to be defined. A natural choice [4] is the mean-square displacement between two configurations a and b, δ_{ab}^2:

$$\delta_{ab}^2 = \min \frac{1}{N} \sum_{i=1}^{N} |\mathbf{x}_i^a - \mathbf{x}_i^b|^2 \tag{3}$$

where $\mathbf{x}_i^{a(b)}$ denotes the position of monomer i in system $a(b)$. The minimization is to be performed over translations, rotations and reflections.

With δ_0^2 denoting the distance to the ground state, and $P(\delta_0^2)$ the corresponding probability distribution, we can define the probability for the system to be found in the vicinity of the ground state as:

$$p_0 = \int_0^{0.04} P(\delta_0^2) d\delta_0^2 \tag{4}$$

The folding temperature T_f is then defined as the temperature where $p_0 = 1/2$.

3 Results

We use six sequences, of length $N = 20$ (Table 1), chosen in [2] to represent a variety of behavior, and examine their kinetic properties for short (intermediate) times. That is, the times should be large enough so as not depend on details of the MC-method (here the normal Metropolis algorithm) but smaller than the time scales necessary to equilibrate the systems. For each system $\mathcal{O}(100)$ Monte Carlo runs, each consisting of $3.3 \cdot 10^5$ steps were performed. The time averages of the mean-square displacement (eq. 3)

no.	sequence	T_f
1	$BAAA\ AAAB\ AAAA\ BAAB\ AABB$	< 0.15
2	$BAAB\ AAAA\ BABA\ ABAA\ AAAB$	< 0.15
3	$AAAA\ BBAA\ AABA\ ABAA\ ABBA$	0.23
4	$AAAA\ BAAB\ ABAA\ BBAA\ ABAA$	0.22
5	$BAAB\ BAAA\ BBBA\ BABA\ ABAB$	< 0.15
6	$AAAB\ BABB\ ABAB\ BABA\ BABA$	0.15

Table 1: The six sequences studied. The errors in T_f are approximately 0.02.

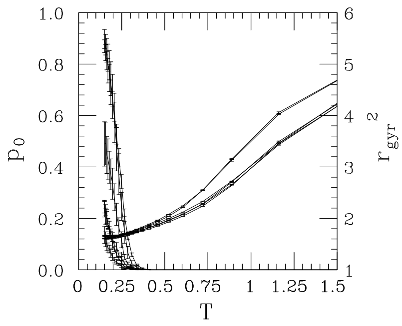

Figure 1: p_0 – the probability for the system to be found close to the native state, and r^2_{gyr} – the radius of gyration squared, as a function of temperature.

$\langle\delta^2\rangle_t$ are obtained by averaging over consecutive subsets of conformations with temporal extension t (this corresponds to an average time separation of $\approx t/3$). The simulations were performed at three different temperatures $T = 0.15, 0.23$ and 0.4. The two lowest correspond to folding temperatures of sequences 3 and 6 respectively (see Table 1), whereas $T = 0.4$ is low enough to correspond to compact states but high enough so that the ground state is almost completely de-populated. This can be seen from Fig. 1 (data from [2]), in which the population of the ground state p_0 and the radius of gyration (squared) is plotted versus temperature.

In Fig. 2 the time dependence of $\langle\delta^2\rangle_t$ is displayed for short times. The data are well parametrized by a behavior of the type

$$\langle\delta^2\rangle_t \propto t^\nu \tag{5}$$

with a temperature dependent exponent ν. For a harmonic chain one has $\nu = 1/2$ while the $T \to \infty$ limit gives $\nu = 1$ [5]. At short times the high temperature limit is valid also for this model. For the sequences here examined ν grows with temperature (Table 2) in a roughly sequence independent manner. At a fixed (absolute) temperature there is a small difference in that the best folder is the chain with the slowest kinetics. This can partly be attributed to the fact that this system spend a lot of time in the vicinity of the native state although this effect mainly should affect the pre factor in eq. 5. On the other

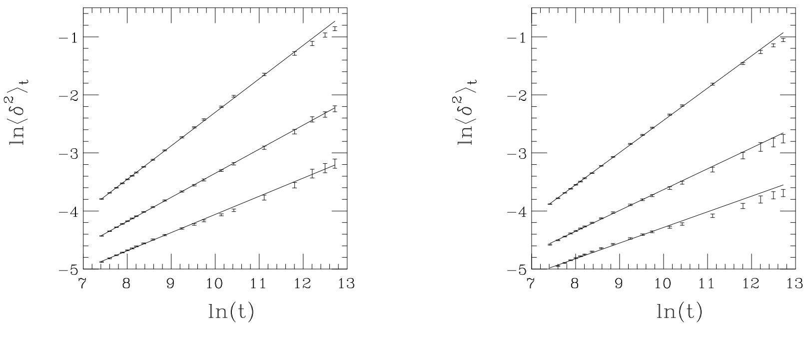

Figure 2: $\langle\delta^2\rangle_t$ as a function of t for sequence no. 1 and 3. The data correspond to, from top to bottom, the temperatures $T = 0.4, 0.23$ and 0.15. The solid lines are linear fits.

hand when comparing the systems at the same "physical" temperature the difference is larger and in the other direction. The best folder now has the fastest kinetics. That is, when comparing the sequences at temperatures chosen such that the chains spend an equal amount of time in the native state, high

no.	T=0.15	0.23	0.4
1	0.31	0.41	0.58
2	0.31	0.40	0.57
3	0.27	0.36	0.56
4	0.29	0.37	0.56
5	0.32	0.42	0.59
6	0.30	0.40	0.58

Table 2: The exponent ν in eq. 5 for the different sequences at different temperatures. The errors are approximately 0.03

thermodynamic stability implies fast kinetics. This means that the crucial requirement to be satisfied for a sequence to represent a good folder is thermodynamic stability of the ground state. The more stable the ground state is the faster becomes the kinetics, at the folding temperature, towards this ground state. This is in line with what was found in [3]. With a somewhat different distance measure a similar study was performed for the IMP-model in [5]. The investigation was here focused on the behavior of the exponent ν at fixed temperature when the strength ϵ of the quenched disorder was changed. The results ranged from $\nu \approx 2/3$ for the ordered system ($\epsilon = 0$) to $\nu \approx 1/2$ for a chain with strong quenched disorder ($\epsilon = 10$). The difference between different realizations of the couplings η_{ij} at fixed value ϵ was found to be quite small although no distinction was made between these realizations in terms of their thermodynamic properties.

Next we study how the structural fluctuations decay as the temperature is lowered. All thermal averages where obtained with the "simulated tempering" method [6]. The torsional angles are defined by $\cos\phi_k = (\hat{b}_k \times \hat{b}_{k+1}) \cdot (\hat{b}_{k+1} \times \hat{b}_{k+2})$. In Fig. 3 the fluctuation in these angles, defined by $\langle \phi_k^2 \rangle - \langle |\phi_k| \rangle^2$ is showed as function of temperature. The lines represent a decreasing sequence of temperatures ranging from $T = 1.67$ (top) to $T = 0.15$ (bottom). As can be seen, the sequence with good folding properties (seq. 3)

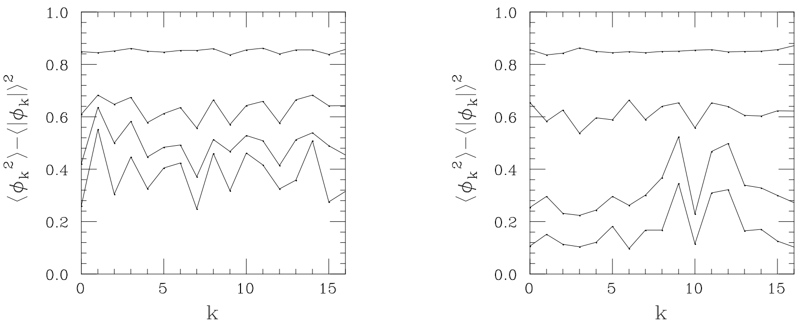

Figure 3: Fluctuation of torsional angels for sequences no. 1 and 3. The data corresponds to, from top to bottom, the temperatures $T = 1.67, 0.41, 0.23$ and 0.15.

have a more drastic "freeze out" of the (torsional) degrees of freedom, and at the lowest temperature, where $p_0 \approx 0.9$ only a few (3) of the torsion angles have significant fluctuations. For sequence 1 on the other hand $p_0(T = 0.15) \approx 0.25$, *i.e.* the thermodynamic stability requirement for the ground state is not yet satisfied, all the angles show rather large thermal fluctuations. These measurements do however not provide us with any information concerning the amount of correlations present in this process. In order to examine this issue we estimate an effective number of degrees of freedom for the chain, by calculating the

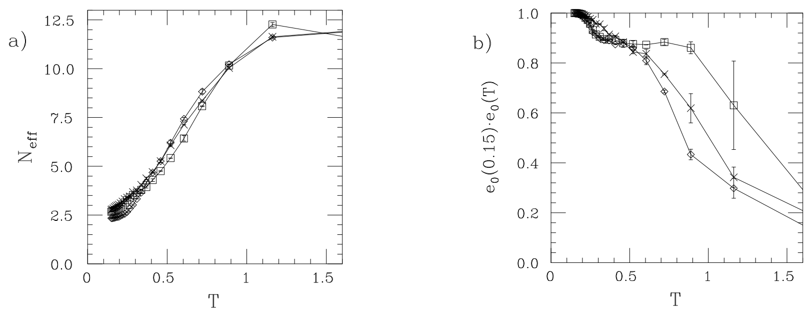

Figure 4: **a)** The "effective size" of the chain as a function of temperature. **b)** The scalar product of the leading eigenvector at T with the one at $T = 0.15$. The data correspond to seq. 1 ($\times$), seq. 7 ($\diamond$) and seq. 16 ($\square$) respectively.

eigenvalues to the correlation matrix

$$\rho_{ij} = \langle \hat{b}_i \cdot \hat{b}_j \rangle \tag{6}$$

We define an "effective size" of the chain as

$$N_{eff} = \frac{\mathrm{Tr}\ \rho}{\lambda_0} \tag{7}$$

where λ_0 is the largest eigenvalue of ρ. For this particular choice of correlation matrix the trace equals $N - 1$. As can be seen from Fig. 4 a there is a gradual decrease in the number of effective degrees of freedom as the temperature is lowered. This is in contrast to the behavior expected from a quadratic potential, for which N_{eff} would have been temperature independent. Thus the energy landscape becomes more and more asymmetric as the temperature goes down. Comparing with the behavior of r^2_{gyr} and p_0 in Fig. 1 we see that this "freezing out" effect occurs mainly in the compact state but before the ground state is populated. In order to see how high in temperature the asymmetry defined at the lowest temperature persists we show in Fig. 4 b the scalar product between the dominant (normalized) eigenvector at $T = 0.15$, $e_0(0.15)$ with the corresponding one at temperature T, $e_0(T)$. This represents a measure on how similar the direction specifying the largest structural fluctuations is compared with the principal direction of the "valley" hosting the native state, or rather the dominating "valley" at $T = 0.15$. For seq. 3 and 4 at least these coincide. As can be expected at high temperatures there is no memory of this direction and although the qualitative behavior is similar for the different chains the sequence dependence is rather large compared to that of N_{eff}. This reflects the fact that $e_0(0.15) \cdot e_0(T)$ is directly related to the properties of the low-end of the energy spectrum while N_{eff} is more general in character.

4 Conclusions

We have studied the kinetic behavior of a 3D off-lattice protein model in terms of shape diffusion at short time scales in connection with folding. We find that sequences with very different thermodynamic behavior, e.g. with respect to the stability of the ground state, have fairly similar kinetic behavior at short time scales. This suggests that the crucial requirements for a chain to have good folding properties are mainly thermodynamic in nature. Furthermore, we examine the behavior of the structural fluctuations of the system – how large are the correlations in the process of freezing out the configurational fluctuations? By estimating an effective number of degrees of freedom present in the system we find that the collective effects in this process are indeed very high and that the asymmetry of the energy landscape is increased as the temperature is lowered.

References

[1] For a review, see M. Karplus and A. Šali, "Theoretical Studies of Protein Folding and Unfolding", *Curr. Opin. Struct. Biol.* **5**, 58 (1995).

[2] A. Irbäck, C. Peterson, F. Potthast and O. Sommelius "Local Interactions and Protein Folding: A 3D Off-Lattice Approach", to appear in *J. Chem. Phys.*

[3] A. Irbäck, C. Peterson and F. Potthast "Identification if Amino Acid Sequences with Good Folding Properties in an Off-Lattice Model", *Phys. Rev.* **E 55**, 860 (1997).

[4] G. Iori, E. Marinari and G. Parisi "Random Self-interacting Chains: a Mechanism for Protein Folding", *J. Phys. A.* **24**, 5349 (1991).

[5] P. Pliszka and E. Marinari "On Heteropolymer Shape Dynamics", *Europhys. Lett.* **22**, 3 (1993).

[6] E. Marinari and G. Parisi "Simulated Tempering: A New Monte Carlo Scheme", *Europhys. Lett.* **19**, 451 (1992).