NORDITA preprint

CERN LIBRARIES, GENEVA

NORDITA - 98/20 CM

SCAN-9811011

5W9845

CHARACTERIZATION OF HETEROPOLYMER COLLAPSE

Ø. Skjæveland and K. Sneppen.

Nordita, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark.

Submitted to Phys. Rev. Lett.

NORDITA · Nordisk Institut for Teoretisk Fysik

Blegdamsvej 17 DK-2100 København Ø Danmark

Characterization of Heteropolymer Collapse

Oyvind Skjæveland and Kim Sneppen

NORDITA,

Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

23 March 1998

Abstract

This paper proposes a method to enumerate and characterize the multi dimensional rearrangement that a heteropolymer collapse represents. We consider collapse of model polymers using Langevin dynamics, and analyze folding pathways by measuring how matches to the ground state develop, when viewed from each monomer in the polymer. The method is illustrated by following one specific heteropolymer, which was selected to have repeatable dynamical access to its ground state. For this heteropolymer we found evidence of a nucleation event that precedes the correctly folded state.

Polymer chains are interesting because proteins are formed by folded polymers. Simplified models for the study of thermodynamics of polymer chains with two types of monomers have been done by using lattice [1, 2] and off-lattice simulations [3] both studied by Metropolis algorithms. Approaches to study dynamics in the form of molecular dynamics [4, 5] or Langevin dynamics have been suggested by [5]. The benefit of molecular or Langevin dynamics is that it makes it possible to address questions related to dynamics directly, whereas the benefit of Langevin dynamics in particular is that each amino acid experiences such a viscous environment that motion on the scale of a single amino acid is overdamped. Thus Langevin dynamics opens for more efficient simulation of polymer chains, where the single timesteps can be large, in our case $dt \sim 0.25\text{psec}$, and thus dynamics can be followed for many microseconds.

In the present paper we study folding of polymers by Langevin dynamics. In particular we investigate the dynamics of the folding process and develop a method to examine folding pathways of polymer chains.

We model a polymer as $i = 1, 2, \dots, n$ monomers on a chain. For the overdamped dynamics under consideration the Langevin dynamics for the monomers can be expressed by the updating rule:

$$d\mathbf{r}_i = -\frac{1}{\gamma} \frac{dV_i}{d\mathbf{r}_i} dt + \sqrt{\frac{2kT}{\gamma}} dt \eta_i(\mathbf{r}, t) \quad (1)$$

where $\mathbf{r}_i$ is the coordinates of monomer i and η_i is Gaussian white noise with zero mean and variance 1. The noise term represents the kicks from the water

molecules that are not explicitly included in the model. Its amplitude is given by the fluctuation dissipation theorem with $kT/\gamma = D$ where the diffusion constant D of a single amino acid is about $10^{-9}\text{m}^2/\text{s}$ [6] for $T = 300\text{K}$. If we have a mass M of an amino acid which is 75 times the mass of the proton (equal to mass of glycine) one obtains $M/\gamma = 0.03\text{psec}$. This is a short timescale compared to characteristic motion of an amino acid, which justifies our use of the overdamped dynamics in eq. 1, see also [5]. The total force on each residue i in the polymer chain is a sum of 3 terms: a very strong covalent bond between nearest neighbors on the chain, a Lennard-Jones (LJ) potential between any residues that are not immediate neighbors, and a bond angle potential that tends to keep angles between covalent bonds along the chain close to 120 degrees. The attractive part of the LJ potential will depend on the hydrophobicity of the involved residues, and we will choose its strength to be 1 if both hydrophobic, 1/2 if only one hydrophobic, and zero if neither are hydrophobic (relative to the values listed in table 1). All potentials are selected to fit typical values of free energies estimated from properties of real amino acids, obtained from [8]. In table 1 we have listed the potentials used in the simulations.

To be able to compare two different realizations of the same polymer we introduce a distance matrix (DM). For both realizations that should be compared we for each monomer $i = 2, \dots, n-1$ do the following steps (In a Cartesian coordinate system):

Step one: Move monomer i to the origin.

Step two: Rotate the structure around the origin to place monomer $i-1$ on the negative x-axis and monomer $i+1$ in the x-y plane, positive y coordinate.

Step three: Define matrix element D_{ij} , $j = 1, \dots, n$ as the distance between monomer number j in the two structures.

Finally, because all monomers and forces in our simple polymer model are mirror symmetric, nothing prevents the polymer from folding into mirror symmetric states, and we want to identify these as being the same state. Thus as a step four we reflect the polymer in the plane formed by monomers $i-1, i, i+1$, compute the D'_{ij} s for both comparisons, and take the one which have the minimal sum $\sum_{ij} D_{ij}$.

The resulting matrix D_{ij} measures all monomer distances as observed from each monomer. The matrix D_{ij} has value zero on the diagonal. Typically, if the polymer is in a random coil state, it tends to have increasing values when one looks at more off diagonal values. Further, for each i , the sum $d_i = \sum_j D_{ij}$ defines a metric. Furthermore, $d = \sum_i d_i$ is also a metric. A region of the matrix where D_{ij} is close to zero will signal a close match between corresponding regions of the compared structures.

To analyze folding pathways of a given polymer, we need to define the states that the polymer visits during its history:

1) To reduce the influence of noise we quench the temperature to zero and sim-

ulate until a local minima is reached.

2) Define a state as a collection of minima that have small distances $d = \sum_{ij} D_{ij} < d_{min}$ to a reference structure. $d_{min} > 0$ represents a coarse graining over local minima that deviate only marginally from each other.

3) For a given ensemble of transitions the states are class divided into basins: A state A belong to a basin α if there is a transition from a state B_α in basin α to A and a transition from state A to a state C_α in basin α . In a basin, the state which has the lowest potential energy is called the basin bottom.

Basins have been suggested earlier by Berry et. al. [4], our way to define them deviates by saying that they represent kinetically equivalent states for the runs under consideration. Notice that by our definition then an increased number of folding histories or a higher temperature typically will lead to larger basins of equivalent states.

We now consider heteropolymers made of two types of monomers, hydrophobic and hydrophilic. Most of the heteropolymers we have analyzed do not reach well defined accessible ground states. In order to find a heteropolymer which repeatedly reaches the same well defined ground state from different initial configurations one needs to do a careful search. Inspecting a few hundred binary heteropolymers of length $n = 20$ we have located a single sequence which repeatedly, for different initial configurations, folds into the same low energy state on reasonable timescales. The obtained good folder has the sequence 00000000111110000100 where 1 indicates a hydrophobic and 0 a hydrophilic monomer.

We analyze the folding history of the selected polymer. In figure 1 we show the angular and hydrophobic energies for the visited basin bottoms as function of time. After a short time $t \approx 15\text{nsec}$, not visible on the scale of the figure, the system has collapsed to its compact state. After a time of about $t \sim 50\mu\text{sec}$ it reaches its ground state. When inspecting angle potential separately, one observes that the overall guiding in the compact state is done through subsequent lowering of the angle potentials, whereas the hydrophobic potential remains roughly constant after the fast initial collapse.

An important question related to heteropolymer dynamics, is the nature of protein folding. Does this take place as a random event in a disordered energy landscape suggested by studies of lattice models, or due to some guiding principle. Some guiding seems necessary to circumvent Levinthal's paradox [11], and has indeed been proposed under many names: in the form of a folding funnel [12, 13], in the form of nucleation with subsequent addition of structures as suggested in experiments of folding of Barnase [9] and in the form of hierarchical assembly of substructures [10]. In our selected heteropolymer, some guiding seems to be present in the form of subsequent lowering of angular potentials.

To look for pathways we have compared folding histories of visited basins starting the polymer from different initial configurations. We find that the histo-

ries are different in the sense that the basins visited in one history have small or no overlap with basins visited in another history. Thus the folding did not occur as a well defined sequence of events. This somewhat resembles the picture obtained from lattice models, where the folding funnel concept was invented to indicate that folding was governed by weak guiding in a otherwise messy potential.

However as we shall see, this picture is only true when the folding polymer is far from its ground state. The last stage of the folding process has much more structure, as will be revealed through the study of the evolution of individual matrix elements D_{ij} where the folding polymer is compared to its ground state.

Figure 2 displays an example of the matrix where a late stage of the folding heteropolymer is compared to the ground state. The right part of the figure displays the two states of the polymer. We mark close matches, defined as $D_{ij} < 2.5\text{\AA}$, with black. We observe regions of close match in f.ex. monomers 14-20, whereas in particular monomers 2, 6, 10 and 13 are misplaced compared to the ground state as seen from the corresponding columns.

In figure 3 we show three time sequences where we count the number T_i of elements D_{ij} that represent monomers close to the ground state viewed from monomer i . Referring to figure 2, T_i is then the number of black squares in row i . The first row in figure 3 shows the i at which the fit is maximal at each time. The second row shows matches measured from $i = 10$, and in the third we count matches measured from monomer $i = 15$. One observes that stable matches first develop late in the folding history. There is essentially no permanent structure developing before $t = 46\mu\text{sec}$. After this a nucleation event around monomer 10 initiates a fast growth of match to the ground state.

In Figure 4 we follow the late stages of the above history in addition to two new folding histories of the same polymer, started from different initial configurations. In all cases we see that structure develops around monomer number $i = 10$ before final state is reached. There is a caveat, however, in the history in the third row, nucleation leads not directly to a match with the ground state, but instead to a match with the state shown in figure 2. The nucleation to this intermediate state may be inspected in the last row of figure 3. This state has easy direct access to the ground state, and it transforms to the ground state after about $1\mu\text{sec}$. We note that in all cases the size of the nucleating region is about 10-15, which is rather close to the total system size of 20. Accordingly there is not much folding history to follow after the correct nucleation event for our very limited heteropolymer size of 20.

In conclusion, we have described the folding history of a heteropolymer that repeatedly folds to a specific state. After a fast initial collapse driven by hydrophobic forces, we observed rearrangement at constant hydrophobic potential but stepwise lowering of the angular potential. During this process there was no specificity in the sequence of states visited. We found that the ground state was reached shortly after a nucleation event was initiated. The nucleation did not start from a specific global state of the polymer, but from a specific site located

Potentials:	form:	constants:
Lennard Jones	$E_0 \left(\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right)$	$E_0 = 572 pN\text{\AA}, \quad r_0 = 4.45\text{\AA}$
Covalent bonds	$k_c(r_i - r_{i-1} - r_c)^2$	$k_c = 290 pN/\text{\AA}, \quad r_c = 5\text{\AA}$
Angle potential	$k_a(\theta - \theta_0)^2$	$k_a = 166 pN\text{\AA}/\text{rad}^2, \quad \theta_0 = 120^\circ$

Table 1: Covalent bonds act between adjacent monomers along the polymer. Angle potentials act between adjacent bonds along the polymer. The Lennard Jones potential has its minimal value at $r_{min} = 5\text{\AA}$, and to avoid numerical instabilities it is replaced by a harmonic form for $r < 4.2\text{\AA}$. The total potential V is the sum of the three terms.

in its hydrophobic middle.

As a perspective we would like to stress that by going from a single number to a matrix in characterizing differences between polymers and their folding pathways, we open for discussion of structure comparisons on all levels. The present study has demonstrated how folding can be characterized by this matrix, in spite of the limited history that could be assigned to our very simplified heteropolymer model. For a larger polymer, i.e. a numerical study of a protein like Barnase where experiments suggest a specific folding history [9], the distance matrix would capture the formation of early substructures as well as how these are aggregated to get the native state. Finally by focusing on the particular parts of the distance matrix that are associated to the monomers that make up the functional site of the protein one can define a metric that associates to degree of functionality.

In summary, the distance matrix provides a way to characterize folding which is quantitative but still relates directly to the qualitative pictures normally used when describing configurational changes in proteins.

References

- [1] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas & H. S. Chan, (1995) *Protein Science* **4**, 561.
- [2] H. Li, C. Tang, & N. S. Wingren, (1997) NEC preprint.
- [3] A. Irback and F. Potthast, (1995) *Journal of Chemical Physics* **103**, 10298.
- [4] R.S. Berry et. al. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 9520.
- [5] D.K. Klimov and D. Tirumalai, (1997) *Phys. Rev. Lett.* **79**, 317
- [6] S. Vogel, (1988) *Life's Devices-the physical world of animals and plants*. Princeton University Press, New Jersey, U.S.A.
- [7] A. Irback, C. Petterson and F. Potthast, 1997, *Phys. Rev. E* **55**, 860.
- [8] A. Wallqvist and M. Ullner, 1994, *PROTEINS* **18** 267.
- [9] B. Nolting, R. Golbik, J. L. Neira, A. S. Soler-Gonzales, G. Schreiber and A. R. Fersht, *Proc. Natl. Acad. Sci.* **94** 826 (1997).
- [10] A. Hansen, M. H. Jensen, K. Sneppen, & G. Zocchi, (1997) *Physica A* in press.
- [11] C. Levinthal, (1968) *J. de Chimie Physique et du Physico-Chimie Biologique* **65**, 44-45.
- [12] H. Frauenfelder & P. G. Wolynes, (1994) *Physics Today* **47** 58.
- [13] P. G. Wolynes, J. N. Onucic & D. Tirumalai, (1995). *Science* **267** 1619.

Figure 1: A folding history for the selected heteropolymer. The upper curve shows Lennard-Jones (Hydrophobic) energy, and the lower curve shows angular energy, both measured by excess energy relative to the ground state. On the scale of the figure, bond potential energies remain constant. After a time of about 15nsec both radius of gyration and hydrophobic potential remains roughly constant. The folding history is however characterized by a stepwise lowering of the angular potential. Simulations with different initial configurations give similar patterns with a stepwise lowering of angular potentials before final state is reached. All simulations are done slightly below the melting transition temperature.

Figure 2: Two structures and the matrix D_{ij} that describes their differences. Dark parts of the chain are hydrophobic, light parts hydrophilic. Both structures are minima on the energy landscape and the lowest structure is the global minimum. Values of the matrix that are less than $2.5\text{\AA}$ are colored black, the other white. From the four nearly white columns one observe that elements $k=2,6,10$ and 13 are misplaced, as can indeed be seen in the structures. All rows which define orientation or position using any of these elements will thus be mostly white. The bottom right black rectangle of the matrix tells that elements $14-20$ have same conformation in both structures. The displayed structures are in fact historically related, the upper structure switches to the lower in about $1\mu\text{sec}$.

Figure 3: Display of the number of elements that are close to their positions in the ground state for the folding history displayed in Figure 1. The number is measured from different local matches between the folding polymer and its ground state. Distance is measured by D_{ij} , and for each fit i the T_i count the number of elements which have $D_{ij} < 2.5\text{\AA}$. In the left figure we show $T_{max} = \max(T_i)$ over all fits between the developing and the final state, in the middle figure T_{10} counts the number of close monomers when monomer number 10 is used as reference, and in the right figure we show T_{15} . At early times $T_{15} = T_{max}$ which reflects that match have developed around monomer 15. Later, at time $t_n = 45\mu\text{sec}$ this structure dissolves to allow for a match that develops around monomer $i = 10$. At that time a structure nucleate around monomer number 10, which fast spreads throughout the polymer to leave it in its ground state at time $t_n + 0.5\mu\text{sec}$.

Figure 4: As in figure 3 except we here focus on the late time of structure development. Further we show the result of three histories obtained by starting at three different initial configurations of the heteropolymer. History a and b are similar, folding history number c differs by first sight significantly. However notice that the state with $T_3 = 16$ is close to the ground state, see also figure 2. Further if we follow the movement towards state $T_3 = 16$ by using fits between the folding polymer and this particular state we in the last row (c') observe that the folding again resembles a nucleation around monomer number 10.