Kinetic Monte Carlo simulations of protein folding and unfolding

Dmitrii E. Makarov* and Horia Metiu**

*Department of Chemistry and Biochemistry,
University of Texas at Austin, Austin TX 78712
**Department of Chemistry and Biochemistry and Department of Physics,
University of California, Santa Barbara CA 93106

ABSTRACT

We report here kinetic Monte Carlo (KMC) studies of two problems: (1) Unfolding of the protein titin under mechanical tension and (2) The formation of native contacts in the process of folding of small, single domain proteins. Both cases involve long time scales and cannot be directly addressed by molecular dynamics methods. In order to overcome this time scale problem we use a kinetic Monte Carlo method, in which elementary processes (such as the formation of a single native contact or dissociation of a hydrogen bond) are viewed as first order processes whose rates are calculated from a crude microscopic model. Our simulations explain the observed behavior in unfolding of titin when pulled by a force and the correlation between the folding rate of single domain protein and the number of contacts in the folded state.

Keywords: protein folding, AFM, titin.

1 INTRODUCTION

Recent advances in single molecule manipulation techniques (such as atomic force microscopy) allowed researchers to stretch single protein chains by applying forces at their ends. The experiments measure the connection between the pulling force and the length of the protein. The data reveals the force under which the folded domains in the protein fall apart and also gives information regarding the elasticity of the polypeptide chain.

A second problem we study is the folding rate of simple single domain protein, in an attempt to explain why the folding rate depends on the number of contacts in the native (folded) state and what the nature of this dependence is.

The kinetics in both problems takes place on a time scale that is too long for molecular dynamics calculations. Because of this we treat them by performing kinetic Monte Carlo simulations on models that attempt to represent the physics of the problem semi-quantitatively. We use microscopic models to calculate the rate constants of the elementary processes involved in protein folding and unfolding, then use these rate constants as input in a kinetic Monte Carlo algorithm that is capable of dealing with long time scales. By using this approach, we have been able to directly compare our simulation results with experiment.

.

2 TITIN UNFOLDING

Several groups have been able to pull a single titin molecule and measure the dependence of the force on the molecule's extension[1-4]. The AFM pulling experiment performed on a single titin molecule is schematically depicted in Fig. 1.

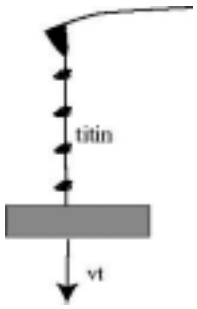

Figure 1. An AFM titin pulling experiment

The titin molecule is attached to a substrate at one end and to a cantilever at the other. The substrate is moved, as indicated in Fig. 1, at a constant velocity v. A typical dependence of the force on time[3,4] is schematically shown in Fig. 2.

Figure 2. Force vs. time in a titin pulling experiment

The titin molecule consists of immunoglobulin-like domains (Ig) arranged in a sequence along the chain[3]. Every time a domain is unfolded under the mechanical load, the chain becomes longer and the force drops exerted on the cantilever drops: The saw teeth in Fig. 2 indicate sequential unfolding of the Ig domains. The peak force, at

which unfolding takes place, was found to be a random variable.

Molecular dynamics studies performed by the Schulten group[5,6] indicate that unfolding of an Ig domain requires breaking of six hydrogen bonds, as shown in Fig. 3.

Figure 3. An Ig27 domain. The numbered lines show the approximate positions of the hydrogen bonds holding the domain together

In order to model the unfolding of a single Ig domain in the titin molecule, we have assumed[7] that each of the six hydrogen bonds in the domain can be described by a double-well potential, the two wells representing the intact and broken bond. The rate constants of breaking and reforming a hydrogen bond are given by

$$k_b(f) = \nu \exp[-U_b(f)/k_BT]$$
$$k_r(f) = \nu \exp[-U_r(f)/k_BT] \qquad (1)$$

where $U_b(f)$ and $U_r(f)$ are the force f dependent barriers to breaking and unfolding. The kinetic Monte Carlo method generates bond breaking/formation events with the probabilities proportional to the rate constants defined by Eqs. (1). We note that while we use the name hydrogen bond, the model works regardless what the nature of bonding is. The domain breaks as soon as the last hydrogen bond is broken. If n is the number of intact bonds in the Ig domain, then n(t) is a random walk that is terminated (the domain is ruptured) as soon as n becomes zero. A typical dependence n(t), generated by our kinetic Monte Carlo simulation, is shown in Fig. 4.

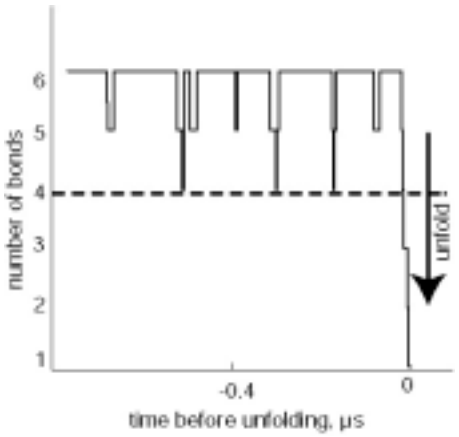

Figure 4. A KMC simulation of a domain unfolding.

This, and many other similar figures, suggest that the domain unfolds irreversibly when the number n of bonds drops below a certain critical value. This behavior results from the fact that bonds act in parallel and the force acting on each bond depends on the number of bonds present, f = F/n, where F is the total force applied to the chain. Thus f increases each time a bond is broken.

The probability distribution of the unfolding force, computed by KMC, is shown in Fig. 5 for a pulling speed of 1 μm/s

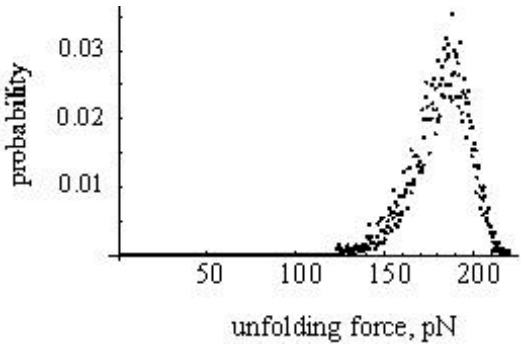

Figure 5. The unfolding force distribution for Ig domains

Force induced unfolding of a titin molecule that includes N Ig domains was simulated using a "bootstrapping procedure" as follows: Let $\sigma(t)$ be the probability that a domain does not unfold up to the time t. Then, for a chain of N domains, the survival probability (i.e., the probability that none of them unfolds) is

$$S(t) = \sigma(t)^N \qquad (2)$$

and the probability distribution for the unfolding time

$$p(t)\,dt = -S'(t)\,dt = -N\,\sigma(t)^{N-1}\,\sigma'(t)\,dt \qquad (3)$$

Thus it is sufficient to know the survival probability $\sigma(t)$ for a single domain to simulate the kinetics of the entire titin molecule. An interesting implication of Eq. 3 is that the

unfolding time (and therefore force) distribution depends on the number of the domains present. Indeed, our simulations explain the fact that the average unfolding forces where found lower for short recombinant titin fragments (with N= 4-8)[3] compared to those in the titin molecule[8].

3 KMC SIMULATIONS OF PROTEIN FOLDING

Plaxco, Simons, and Baker[9] made the intriguing observation that the observed rate constant of folding of small, single domain proteins is correlated with the topology of the native state, quantified in terms of its "contact order". To understand the origins of this behavior, we have studied a kinetic model, in which protein folding is viewed as formation of native contacts, as shown in Fig. 6.

Figure 6. Native contact formation in a chain

Following refs.[9,10], we define a contact in a protein as two residues being within a contact radius *a* from one another and being separated by more than C residues along the chain. A typical choice of these two parameters is *a* = 6Å, C = 12, the results being only weakly dependent on them.

To study the kinetics of the model, one needs a way of calculating the rate constants of contact formation and dissociation. If one uses the Gaussian chain approximation to estimate these, then the rate constant for diffusion controlled formation of a contact between residues i and j is given by[11]:

$$k_{ij}^{+} = \frac{}{\langle \; \rangle} \qquad (4)$$

Here D is the diffusion coefficient and $\langle r_{ij}^2 \rangle$ is the mean square distance between residues i and j. The latter depends on the presence of other contacts in the chain. For example, forming contact *a* in Fig. 6 shortens the loop required to form the contact c.

A contact that has been formed can dissociate again. The rate of dissociation in our model is given by:

$$k_{ij}^{-} = \nu \exp(\beta \varepsilon_{ij}) \qquad (5)$$

where ε_{ij} is the contact binding energy.

Fig. 7 shows a typical result of simulations, the folding time probability distribution for a model chain that forms the contacts {{5,37}, {25, 51}, {17, 41}, {23, 13}, {57, 45}, {1, 33}, {9, 29}}. Except for short times, the observed kinetics are close to exponential[12], which is consistent with the experimental observation that small, single domain proteins exhibit first-order folding kinetics[9,10].

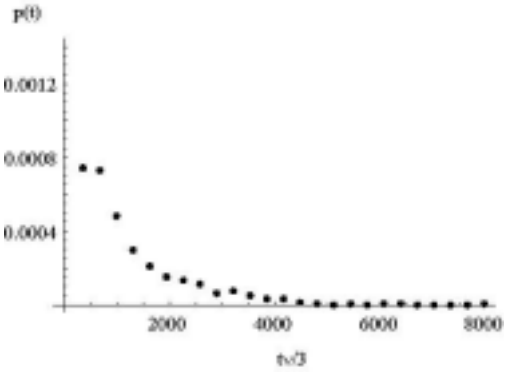

Figure 7. Probability distribution of the folding time for a gaussian chain with N=7 contacts.

We further found that many features of our model are reproduced within a mean-field-type approximation, in which one does not distinguish among specific contacts but rather tracks their total number m. This leads to a model that has the following properties.

1. The total free energy of a chain with m contacts is given by:

$$F(m) = F_0 + m\,\Delta F,\ m = 1, 2, \ldots, N \qquad (6)$$

   Here ΔF is the free energy cost for the formation of a single contact and F_0 is the excess free energy required to form the first contact.
2. The rate limiting step of folding is the formation of the N-th contact, where N is the total number of native contacts

By employing the above two assumptions, the model can be (approximately) solved analytically[7] to find that the effective folding rate is given by:

$$k_f = N\, k^{-} \exp[-\beta F_0] \exp[-\beta N\, \Delta F] \qquad (7)$$

where k^{-} is the contact dissociation rate (5) assumed here to be the same for all contacts.

Eq. 7 reproduces the folding rates of small single domain proteins well[13]. We fitted the dependence of the experimental folding rates on the number of native contacts N by Eq. 7 for a data set of 24 structurally unrelated small single domain proteins studied in[9,10]. The parameters found from this fit are[13]:

$$\Delta F = 0.144\ k_B T$$

$$k_d \exp[-\beta F_0] = 3828\ s^{-1}$$

For the proteins studied, Eq. 7 predicts their folding times to within an order of magnitude. Further, the contact order as defined by Plaxco, Simons and Baker[9], is nearly proportional to the number of native contacts thus suggesting an explanation to the empirical correlation found by those authors.

Simulations are underway in our groups to understand the microscopic basis of the assumptions, which lead to Eq. 7.

REFERENCES

1 M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, *Science* **276,** 1112 (1997).

2 L. Tskhovrebova, J. A. Trinic, J. A. Sleep, and R. M. Simmons, *Nature* **387,** 308 (1997).

3 M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276,** 1109-1112 (1997).

4 M. Rief, J. M. Fernandez, and H. E. Gaub, *Phys. Rev. Lett.* **81,** 4764 (1998).

5 H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, *Biophysical Journal* **75,** 662 (1998).

6 H. Lu and K. Schulten, *Chemical Physics* **247,** 141 (1999).

7 D. E. Makarov, P. K. Hansma, and H. Metiu, *J. Chem. Physics* **114,** 9663 (2001).

8 M. B. Viani, T. E. Schaffer, G. T. Paloczi, I. Pietrasanta, B. L. Smith, J. B. Thompson, M. Richter, M. Rief, H. E. Gaub, K. W. Plaxco, A. N. Cleland, H. G. Hansma, and P. K. Hansma, *Review of Scientific Instruments* **70,** 4300 (1999).

9 K. W. Plaxco, K. T. Simons, and D. Baker, *J. Mol. Biol.* **277,** 985 (1998).

10 K. W. Plaxco, K. T. Simons, I. Ruczinski, and D. Baker, *Biochemistry* **39,** 11177 (2000).

11 A. Szabo, K. Schulten, and Z. Schulten, *J. Chem. Physics* **72,** 4350 (1980).

12 D. E. Makarov and H. Metiu, *J. Chem. Phys.***,** in press (2001).

13 D. E. Makarov, C. Keller, K. W. Plaxco, and H. Metiu, *Proc. Natl. Acad. Sci. USA***,** in press (2001).