Bond Hierarchy and Protein Folding Speed

J. T. Mettetal
Department of Physics,
*Massachusetts Institute of Technol*ogy, *Cambridge, MA 02139*

I present here the results of a study designed to address issues related to folding speed in proteins. Lattice simulations of a protein folding from a random extended to a 'designed' state were studied as the binding energy of monomers Eij was varied according to the monomers distance on the polymer. These simulations suggest that strengthening native interactions between monomers i and j for large |i-j| helps speed the folding process while strengthening interactions for small |i-j| causes a reduction in folding speed. This supports the 'nucleation' view of protein folding where monomers from different parts of the chain come together early in the process to help guide later folds of the structure.

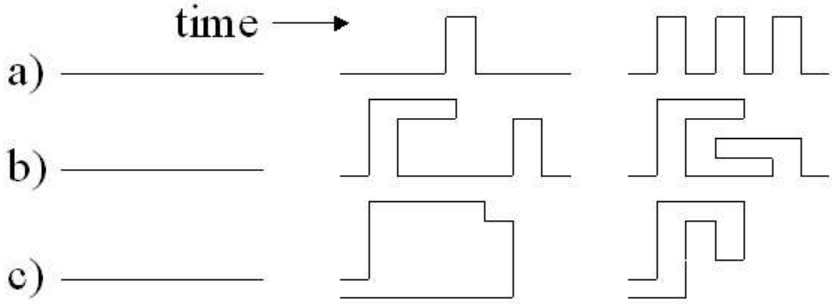
Introduction

Although modern biology has succeeded in describing the mechanisms for many cellular phenomena, its predictive powers would be greatly enhanced if there were a reliable and fast method for predicting binding energies for proteins. To resolve this issue, it becomes essential to first solve the protein-folding problem since these energies typically rely heavily on the precise placement of amino acids. Much work and many papers [1-6] have been devoted to predicting the final structure from the easily determined amino acid sequence. Many important steps forward have been made, but there is still a large need for both computing detailed simulations and fundamental understanding of the mechanisms through which a protein acquires its shape.

One reason protein folding is difficult to understand is that it is a highly entropic phenomenon and the reachable conformation space at any time is large and difficult to describe generally. This means that as a protein progresses from an unfolded to a folded state, it is being driven as much by entropy as by energy. Because the protein is moving toward a less entropic state (there should be only a small region of phase space corresponding to a folded protein) the energy gained must be large enough to compensate for the entropy lost.

Several folding models can be envisioned that give intuitive explanations for this interplay between entropy and energy. In one scenario, monomers nearby on the chain at one point can form native bonds and create a substructure that grows as both sides of the polymer add to this substructure (fig 1a). Similarly, monomers at many points along the chain could form short ranged substructures that subsequently lock together to form a completed structure (fig 1b). Another method could involve monomers at distant locations of the chain forming bonds and limiting the entropy of the resulting subchains, which fold around the resulting 'loose' substructure (fig 1c). The latter is similar to the 'nucleation' theory of folding, which has been suggested as an important method for fast folding [2].

Fig 1



a) Short range bonds form initially and guide formation of other short range bonds
b) Short range bonds form small substructures later locked by long range bonds
c) Long range bonds form first guiding formation of short range bonds

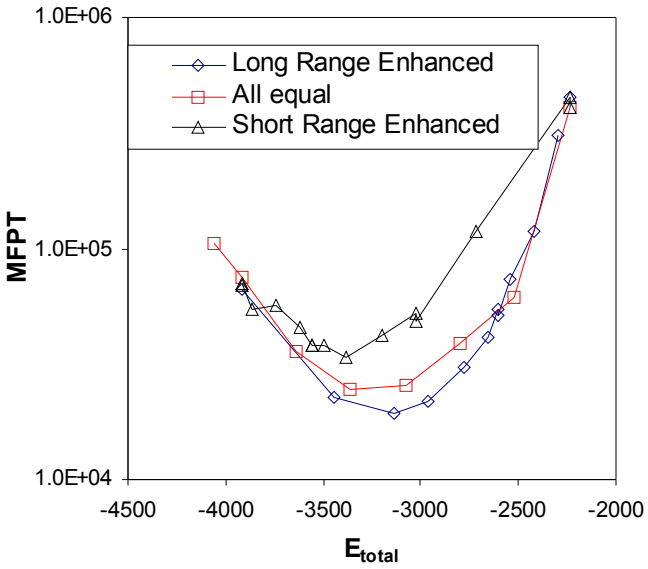
We can address the question of which model explains the folding pattern of proteins going from extended to folded states by examining the folding time of model proteins. Each time interacting monomers meet the decrease in internal energy will cause a bond to form for some period of time. I expect bonds with strong energies to be the ones that fold first and stay folded guiding the folding of bonds that form later. Those with weak bonds can fold and unfold many times only becoming stabilized later in the folding process when there are other interactions to stabilize them. This should mean that by strengthening the bonds that form early and relatively weakening those that form later the protein folding time can be decreased.

Model

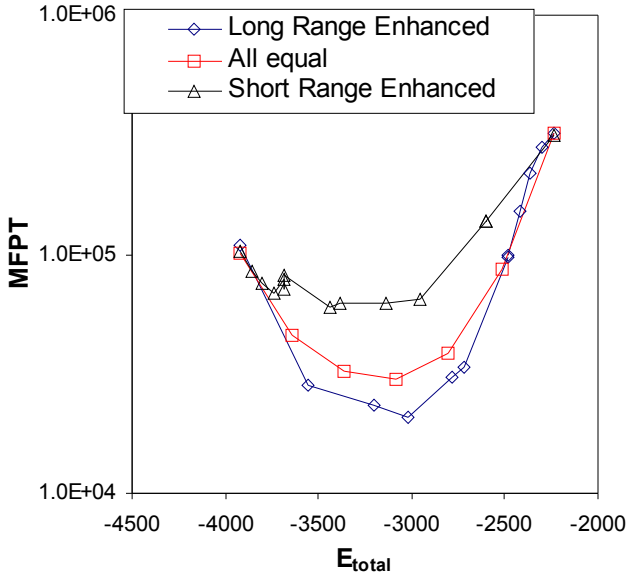
I use a GO lattice model [3] for protein folding identical to that used in previous studies of protein folding [4]. There is evidence to suggest that this model captures many of the dynamic properties of real proteins while having a distinct computational advantage over more complete molecular dynamics models. The protein is modeled as a connected sequence of N nodes on a three dimensional rectangular lattice with each node representing a monomer in the protein sequence. Energies are represented by the value E_{ij} defined as the energy change due to having monomers i and j occupying adjacent lattice sites, and two monomers may also never occupy the same lattice site. Dynamics are given to the molecule through attempting three types of moves: end, flip, and crankshaft. Each attempt is accepted or rejected based on the standard Metropolis rates. A more detailed description of this model is available in the references given below [5]. In this study we confine N=27 which folds to a compact native state on a 3x3x3 cube. Temperature is set to 60, which simply defines the scale E_{ij}/k_BT.

Fig 2

Structure1:



Structure2:



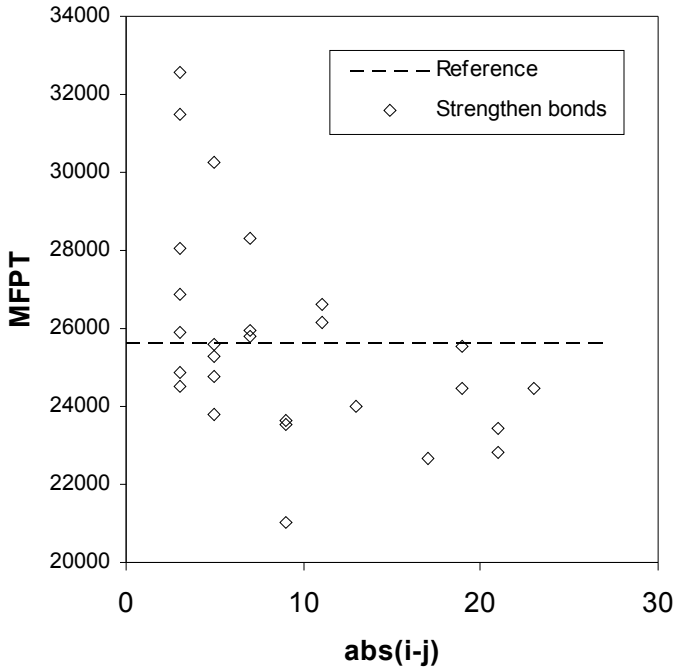
Results

I begin by examining the folding speeds of two random protein structures. It would be more suitable to have examined a larger subset of the possible 3x3x3 structure space, but time has limited this possibility. While folding temperature is held constant, I change the interaction energies for native bonds while keeping non-native bonds held fixed at $E_{ij} = 0$. First, the energies are set to $E_{ij} = -|E|$ for native and 0 for non-native giving an $E_{total} = -27$ E for the native structure. Next, the value of E is varied from –80 to –140 in increments of 10, and Mean First Passage Time (MFPT) from random initial configurations to the native structure is determined from 1000 independent runs. Due to ambiguity in units of E and T, this can also be thought of as holding the energy of the protein fixed while changing the temperature.

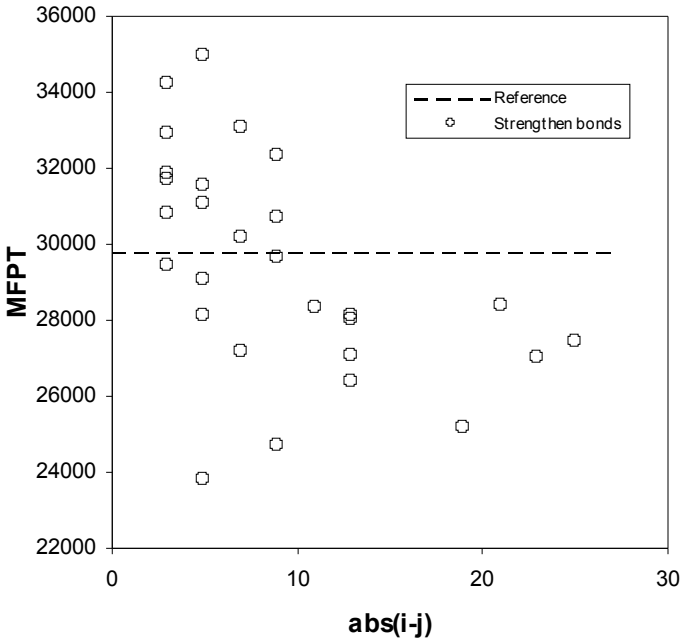
The measured dependence of MFPT on E (fig 2) is low for values of E (E_{total}) near –110 (-2790) and increases as E is changed from this value. The dependence of MFPT on E or E_{total} can be explained through two phenomena. As E decreases, native bonds are becoming weaker so that by the time new bonds form, old ones spontaneously break apart. This limit is the same as a random non-interacting polymer spontaneously forming a compact folded configuration. As E increases, native bonds very rarely break once they have formed, so if they form in a manner that prohibits later structures from forming (knots, blockages, etc.) a longer time must elapse before the bond breaks.

Fig 3

Structure 1:



Structure 2:



To determine which bonds speed protein folding, I next modified the interaction energies in two similar ways. First, I define attractive interactions only for native bonds E_{ij} = -80 for |i-j|>=d and –140 otherwise which corresponds to strengthening short-range bonds and weakening long-range bonds. As d changes the value E_{total} changes as well
based on the number of bonds with |i-j|>=d. In figure 2, I plot the corresponding MFPT for each structure as a function of the E_{total} formed by each value of d. We see that the curve is several times higher (slower folding) than the constant energy curve except on the ends where the two meet when d = 1 or 27 and the E_{ij} models are identical.

To test the other limit, I next define E_{ij} = -80 for |i-j|<=d and –140 which strengthens long-range bonds and weakens short-range bonds. Again as d is varied from 1 to 27 the MFPT is plotted against E_{total} in figure 2, and like the previous case, the two curves meet when d = 1 or 27 as expected. However, now the MFPT is lower (faster folding) than both the constant energy and strengthened short-range model.

Finally, I simulated the folding of both structures with modifications to *single* bonds. Here E_{ij} = 0 for non-native bonds, and E_{ij} = -110 for all native bonds except for a single pair (i', j') where $E_{i'j'}$ = -170. The MFPT for each pair of native bonds i'j' is displayed in figure 3 as a function of polymer distance |i'-j'|. The error on each data point is about 800 time units indicating that the vertical spread is statistically significant. Again, the trend is similar to the previous simulations where strengthening of short bonds causes folding to slow while strengthening of distant bonds causes folding to speed with a crossover distance in this model of |i'-j'| ~ 10 monomers. Although the trend seems to hold strongly, we see that there are

short bonds which, when strengthened, cause a significant decrease in the folding time.

Discussion

We test sample structures for folding time based on a model where native interactions are modified according to the 'polymer distance' |i-j| between the monomers i and j. The evidence presented here strongly indicates that there is an important correlation between bond strength and folding speed based on the polymer distance. This observation should provide implications for which mechanism of protein folding is most correct, and it seems to strengthen the 'folding nucleus' [1] theory as opposed to the 'diffusion-collision' mechanism [6]. It would also be interesting to measure distributions of MFPT to determine how many "slow folding events" occur since variance of Poisson distributed waiting times should scale as (number of events)$^{1/2}$. If this number is much smaller than the number of bonds, it might lead to an interpretation where there are few important slow folding events that later 'guide' the formation of subsequent bonds. Future work in this area should include extension of these results to other folding models to rule out the possibility that these results are an effect of the specific lattice model used here.

References

[1] Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich E I. J Chem Phys (1998) 108:334
[2] Mirny, L.A., Abkevich, V., and Shakhnovich, E.I. PNAS, (1998) v.95
[3] Sosnick TR, Jackson S, Englander SW, DeGradoW. 1996. *Proteins* 24:427–33
[4] Mirny, L.A. and Shakhnovich,E.I. Annu. Rev. Biophys. Biomol. Struct. 2001. 30:361-396.
[5] Pande VS, Grosberg AY, Tanaka T. J Chem Phys (1994) 101:8246
[6] Karplus M, Weaver D. 1976. Proteinfolding dynamics. *Nature* 160:404–6