Protein folding in a reduced Go model

Jaroslaw Labaziewicz*
Center for Ultracold Atoms, MIT, Cambridge, MA 02139
(Dated: May 17, 2007)

Protein folding studies from basic principles present us with a novel perspective on the folding problem. One particular aspect that can be explored with this method is the dependence of folding properties on the alignment and density of attractive interactions between the peptides of the protein. In this paper, we develop a model for protein folding and study the dependence of folding time and folding temperature on the number of native interactions in Go model of protein folding.

I. INTRODUCTION

Protein folding is one of the basic problems at the interface of biology and physics. While we have an extensive knowledge of DNA sequences involved in protein coding, we do not yet understand the relationships between the code and the protein structure, which largely determines the function of the protein. A subset of that larger area of study is the question of the kinetics of protein folding. As noted by Levinthal[1], a protein does not have time to extensively sample the space of spatial configurations in order to fold. A direct conclusion is a necessity for a guiding mechanism to lead the peptide chain from it's unfolded state to it's fully folded state. A number of hypotheses have been used to resolve this paradox. Of particular interest is the claim that local interactions and substructures allow rapid folding of the structure.[2, 3] At the same time, non-local interactions were found to play an important role as well.[3–5] Monte Carlo simulations of folding of a self-avoiding polymer on a cubic lattice[6–10] are well suited to resolving such questions. While they cannot be relied to provide an exact pathway for folding, we expect overall features of the model to translate well to the physical system. In this report, we investigate the dependence of the folding rates and allowed temperature range on the number and position of interactions in a Go-model.

II. MODEL DESIGN

We model the protein as a self-avoiding polymer on a cubic lattice. Consecutive monomers always occupy neighboring sites (manhattan distance). The interactions are pairwise and short-range. As in Go model, we differentiate between native interactions i.e. those active when the protein is in a fully folded state, and non-native interactions. In order to allow for the most efficient structural change, we simulate polymer bending rather than monomer movement, similar to previous work by Scheraga et al.[12]. This allows for

*http://quanta.media.mit.edu/

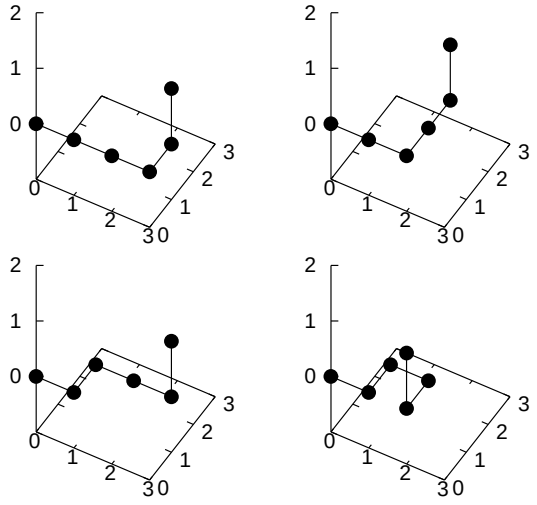

FIG. 1: Legal moves in the model. Starting with a chain (upper-left), we allow bending of any link while maintainling the directions of the remaining links (upper-right). We have found it important to also allow bending the ith and (i+2)th link simultaneously. Two such legal moves are shown in the bottom left and bottom right.

rapid changes in global spatial arrangement in a single move. In a excluded volume model, the choice of moves becomes quite important. The lack of certain transformations can lead to unphysical slow-down in the search for minimum, or even trapping in unfavourable states[11]. In order to alleviate that problem, we allow for both single bend and dual bend at next-consecutive monomers. The dual bend moves are choosen randomly half the time. Example legal moves for a simple chain are shown on Fig. 1.

The model is iterated up to 10^7 times to estimate the folding time. At any iteration, we estimate the energy cost of performing a move. If the energy decreases, the move is always taken. Otherwise, the move is taken with probability $e^{-\beta\Delta E}$. The folding time is defined as the number of iterations required to reach ground state for the first time. We choose our interactions such that the protein ground state is compact. In order to faciliate comparison to literature, we pick a structure equivalent to that of sequence #5 in the work by Shakhnovich et al.[13]

Fig. 2 shows a detail of a single experimental run of our folding algorithm. In the initial state, the protein is randomly arranged in space. After a number of iterations, we arrive at a fully folded state. Close inspection of the energy plot shows that the protein rapidly folds to a compact globule, then slowly "tunnels" to it's final state. This behaviour is consistent with previous results[6].

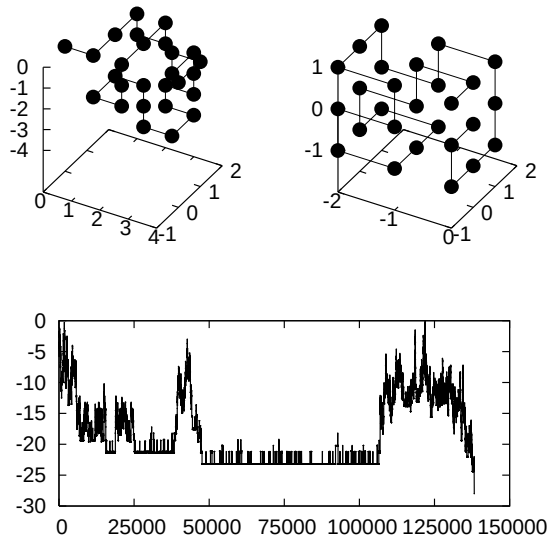

FIG. 2: Protein folding. We start with a randomly arranged protein chain (upper-left) and iterate our model. Eventually, the protein achieves a fully folded state (upper-right). The lower diagram shows protein energy development as a function of time.

III. RESULTS FOR A FULL GO-MODEL

In order to calibrate our model, we first set all native interactions to be strongly attractive, $E_{ij} = -1$, turn off all non-native interactions, and investigate the protein folding time and folding probability as a function of temperature. Then, in order to ascertain that a global hydrophobic behaviour used later will not impede the folding, we add a non-native interaction between every monomer pair at $E_{ij} = -.1$. The resulting plots for average folding time and folding probability are shown on Fig. 3.

We note that the folding time observed in our model is significantly faster to that obtained using only local moves, which we attribute to the flexibility of our movement operations. However, our model is computationally more expensive.

The temperature in which folding can occur forms a very narrow range between .5 and .9. For temperatures lower than that range, we observe freezing of the protein in a local minimum. For temperatures higher then the range, the protein remains in a molten state. We observe very steep increase in folding time away from optimal temperature.

The non-native interactions do not affect folding, indicating that they do not contribute to forming of local minima which can trap the folding process.

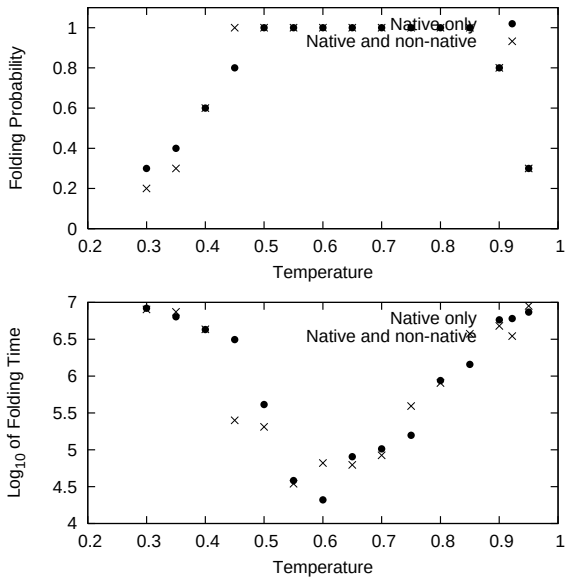

FIG. 3: Go model with all native interactions, and a model with both native and non-native interactions. We see that addition of non-native interactions does not adversly affect folding in Go model.

IV. RESULTS FOR LIMITED GO-MODEL

In subsequent simulations, we limited the number of native interaction while leaving the non-native interactions at $-.1$. We simulate three distinct behaviours. First, we randomly removed 14 of the 28 native interactions. Then, we have removed 9 interactions, all along the edges of the cube. Finally, we have removed 12 interaction, all on the faces of the cube.

The results of the three simulations are shown on Fig. 4. It is clear that removal of a modest number of random native interactions impedes folding substantially. While the decrease of the folding temperature can be expected as the depth of the ground state is decreased, we note that the range of the temperatures defined as $2\frac{T_{max}-T_{min}}{T_{max}+T_{min}}$ is also lowered from .8 to .66. The rate of folding is also substantially lowered, on average by one order of magnitude.

The results for the simulation with native interactions on the edge of the protein removed are much different. While the folding rate is slightly lowered, by about a factor of 2, the range of temperatures over which the protein will fold is increased from .8 to .9, mostly by allowing the protein to fold successfully at lower temperatures. This indicates that the removed interactions contributed to formation of local minima in the folding process.

The results for simulations with native intaractions on the walls of the polymer removed are intermediate between the two others.

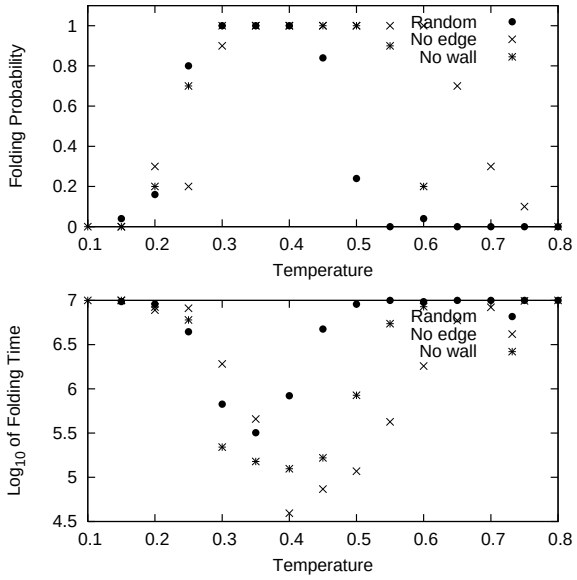

FIG. 4: Limited Go model. The folding times and folding probabilities are indicated for a model with 14 random native interactinos removed, model with 9 interactions on the protein edge removed, and a model with 12 interactions on protein surface removed.

V. CONCLUSION

In conclusion, we developed and investigated a Monte Carlo model of protein folding. Our choice of MC steps allows us to fold the proteins effectively, with $\approx 10^4$ steps required in optimal cases. We have investigated the dependence of folding on the number and arrangement of native interacions, and we found that removal of a half of the interactions can already significantly impede folding. However, by judicious joice of removed interactions, the folding can be made more stable as compared to the full model. These conclusions point out the need to carefully investigate the importance of spatial arrangement of attractive interactions in MC models.

[1] C. Levinthal, J. Chem. Phys **65,** 44 (1968)

[2] J. Bowie, R. Luthy, D. Eisenberg, Science **253,** 164 (1991)

[3] V. I. Abkevich, M. Gutin, E. I. Shakhnovich, J. Mol. Biol. **252,** 460 (1995)

[4] A. R. Dinner, A. Sali, M. Karplus, Proc. Natl. Acad. Sci, **93,** 8356 (1996)

[5] S. Govindarajan, R. A. Goldstein, Biopolymers **36,** 43 (1995)

[6] A. Sali, E. Shakhnovich, M. Karplus, Nature **369,** 248 (1994)

[7] A. Sali, E. Shakhnovich, M. Karplus, J. Mol. Biol. **235,** 1614 (1994)

[8] D. E. Kranbuehl, P. H. Verdier, J. Chem. Phys. **56,** 3145 (1972)

[9] P. H. Verdier, J. Chem. Phys. **59,** 6119 (1973)

[10] N. Go, H. Taketomi, Proc. Natl. Acad. Sci. **75,** 559 (1978)

[11] H. J. Hilhorst, J. M. Deutch, J. Chem. Phys. **67,** 5153 (1975)

[12] Z. Q. Li, H. A. Scheraga, Pro. Natl. Acad. Sci **85,** 6611 (1987)

[13] A. Gutin, A. Sali, V. Abkevich, M. Karplus, E. I. Shakhnovich, J. Chem. Phys. **108,** 6466 (1998)