Simple Lattice Model of Proteins Incorporating Directional Bonding and Structured Solvent

Eamonn M. O'Toole, Rajesh Venkataramani, and Athanassios Z. Panagiotopoulos
School of Chemistry Engineering, Cornell University, Ithaca, NY 14853

Biological catalysts and regulators have assumed major industrial importance, primarily in the production of pharmaceuticals and foodstuffs (King, 1989). Heavy industrial use of proteins requires detailed knowledge of their thermodynamic stability, so that processes can be designed that will not disrupt the biological activity of the proteins. Proteins can be "denatured" with loss of biological activity through nonphysiological conditions such as high temperature, acidic or basic pH, or the presence of some nonaqueous solvents (Cantor and Schimmel, 1980).

A theoretical understanding of protein stability is desirable but presently not complete (Chan and Dill, 1993; King, 1989). We have been studying simple protein models to determine if some features of the experimentally observed thermodynamics of denaturation of real proteins can be captured by such models. The first model we have investigated-the Dill model-represents a protein as a connected series of hydrophobic and hydrophilic beads on a lattice, with nearest-neighbor interactions between the constituent beads (Lau and Dill, 1989). Much work has been done on this and similar simple models of late (Chan and Dill, 1989, 1990, 1991; Miller et al., 1992; Shakhnovich and Gutin, 1989; Shakhnovich et al., 1991; Gutin and Shakhnovich, 1993). These models submit themselves to extensive analysis, at least for short sequences of length 14 to 27. For instance, it has been found that the folding kinetics of 14-bead sequences of the simple Dill model are strongly dependent on the strength of attraction between hydrophobic beads, with intermediate strengths providing the fastest folding through identifiable intermediates (Miller et al., 1992). Shakhnovich and co-workers have examined a model that allows a greater variety of model beads, and have found that only one out of 30 sequences of length 27 fold to the global minimum in a reasonable amount of time (Shakhnovich et al., 1991). They have also found that the lower the energy of the global minimum in their model, then the lower the degeneracy of this state will be (Gutin and Shakhnovich, 1993). In our previous work on the simple Dill model which allows just two model residue types, we have found that the thermal denaturation curve for some sequences of length 48 of this model approximates the sharpness observed for real proteins (O'Toole and Panagiotopoulos, 1992). However, there are some shortcomings in the thermodynamic behavior of this model. In particular, the low-energy states corresponding to "native" states in real proteins are never destabilized at very low temperature, contrary to experiment, and there are no unique low-energy structures for many sequences (O'Toole and Panagiotopoulos, 1993). A key omission from the model is an orientable solvent which interacts with orientable beads on the model backbone in such a way as to make stable conformations in which beads that interact unfavorably with the solvent are oriented away from it and beads that interact favorably with the solvent attempt to surround themselves with solvent. Native protein conformations exhibit this sort of ordering. Also lacking from the model are directed interactions between the model beads that stabilize ordered, regular structures, in much the same way that hydrogen bonding between amino acids-the basic monomer units of real proteins-stabilizes structures such as α-helices and β-sheets in proteins.

From the results we have obtained with the Dill model (O'Toole and Panagiotopoulos, 1992, 1993), it seems that by replacing such orientationally-dependent interactions with interactions that depend solely on the number of nearest-neighbors some important physics is lost. In this note, we describe a model which attempts to include the directional interactions omitted by the Dill model. We also briefly describe the modified Rosenbluth and Rosenbluth and Metropolis methods that we use to study this model. Finally, we give some preliminary results obtained from the model.

Model and Methods

This model is based on previously available lattice models of water (Bell, 1972; Meijer et al., 1981). These models attempt to capture several characteristic properties of real water-directional bonding which depends on the orientation of the solvent molecules, tetrahedral coordination, and the density maximum seen in water at around 4°C. While the original work on these models (Bell, 1972; Meijer et al., 1981, 1982; van Royen and Meijer, 1984, 1986) suggested that they produced results in good qualitative agreement with experimen-

Correspondence concerning this article should be addressed to A. Z. Panagiotopoulos.

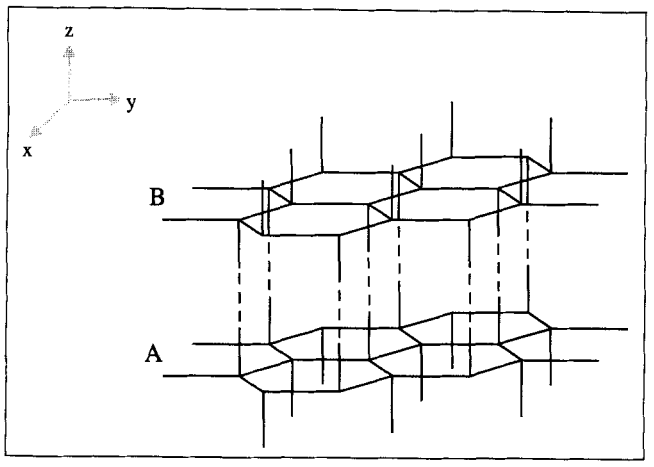

Figure 1. Hexagonal lattice, distorted for clarity.

This lattice consists of layers of hexagons stacked in the z axis so that the hexagons align; individual hexagonal layers are corrugated, but appear flat for clarity. The layers have also been separated; there are hexagonal structures in the x-z and y-z planes, but they do not align. Layer B is the mirror image of layer A.

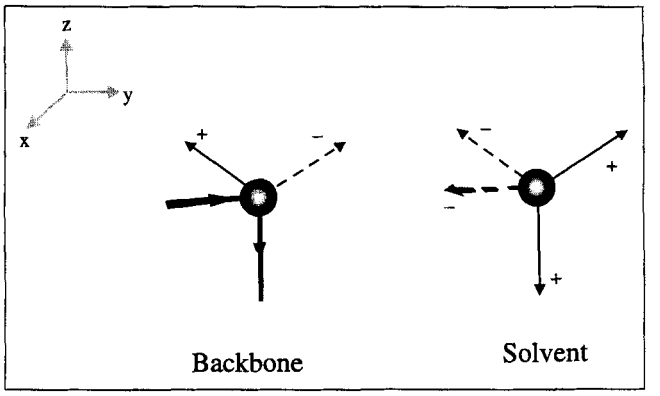

Figure 2. Model backbone bead and a model solvent.

Appendages from both types of bead are in a tetrahedral orientation; leftmost appendages from both beads in the figure, extending in the $-y$ direction, do not lie in the plane of the page. To give a sense of perspective, lines projecting into the plane of the paper diminish in thickness; thick lines through the backbone bead represent the backbone links. The arrows on these links define the direction along the backbone from beginning to end, as described in the text. The positive arm on the backbone bead extends to the left as one looks down the backbone in the direction indicated along the backbone.

tal results for water, more recent work (Borick and Debenedetti, 1993) has shown that the behavior of the low-temperature density maxima does not agree with experiment. This unphysical behavior is attributed to the repulsive interaction in the models responsible for the density maxima. However, the anomalous low-temperature expansion of water is not of interest to us. The models do incorporate a simple directional-bonding mechanism which we may use as the basis for a protein model which will allow for directed bonding between model backbone beads. In addition, the model solvent can assume structure analogous to that seen in water, with the solvent capable of forming a maximum of four directed bonds with its nearest neighbors. The lattice that forms the basis for our model is the 3-D hexagonal lattice, a tetrahedrally-coordinated lattice that allows for helical structures. This lattice is illustrated in Figure 1. Each vertex of this lattice is occupied by either one of two entities: a model backbone bead, or a model solvent, both of which are illustrated in Figure 2. To define orientation, each model bead has "arms" which point towards the vertices of a tetrahedron and which are labeled as positive or negative. Overlap between arms of opposing sign on neighboring beads will result in an interaction as explained below and described in Table 1; if the overlaping arms are of the same sign then there is no interaction.

There are four arms extending from each lattice site on which a solvent is located, as indicated in Figure 2. Two of these arms are designated positive, two are designated negative and there is a favorable interaction when a negative arm on a solvent overlaps with a positive arm on a nearest-neighbor solvent. Each model backbone-bead has two arms, a single positive arm and a single negative arm. There are two types of backbone bead, designated hydrophobic and hydrophilic, with favorable interactions between constituent backbone beads when a positive arm extending from a backbone bead overlaps a negative arm extending from a nearest-neighbor backbone bead. A "bond" formed in this way can stabilize regular model protein structures on the beads which is independent of their orientation similar to the simple Dill model. This attraction is of equal strength to the favorable interaction between two backbone beads when a positive arm on one bead overlaps a negative arm on a nearest-neighbor bead.

Two methods are used to examine the behavior of this directional protein model. The modified Rosenbluth and Rosenbluth (1955) method described in a previous publication (O'Toole and Panagiotopoulos, 1992) has been adapted to study only the vacuum variation of the model described above. Extending the technique to incorporate the solvent is problematic, and has not been attempted. This technique involves generating a sample of protein conformations. Each conformation is generated using a step-by-step growth of the protein on a lattice in multiple-bead sections, with the conformation of each successive section being chosen with probability proportional to its Boltzmann weight. Because of this generation process, every conformation in the sample has an associated statistical weight. Average properties are calculated from the sample and its correcting weights, as described

Table 1. Interactions between the Entities Possible in the Model: Hydrophobic Beads (H), Hydrophilic Beads (P) and Solvent S

Interaction	Strength
$P-P$	-1
$P-H$	-1
$H-H$	-1
$P-S$	-1
$H-S$	$+1$
$S-S$	-1

*A negative sign indicates attraction and a positive sign indicates repulsion; interactions are scaled with the hydrophobic-hydrophobic attraction.

elsewhere (O'Toole and Panagiotopoulos, 1992; Szleifer et al., 1992).

The method of Metropolis et al. (1953) has been applied to the full model incorporating both orientable solvent and directional backbone, and also to the vacuum variation where the orientable solvent is removed. To apply the Metropolis method, we use a set of moves to perturb a starting conformation with each perturbation being accepted or rejected according to the probability given in Metropolis et al. (1953). If a perturbation is accepted, the resulting conformation becomes the basis for all subsequent perturbations until the next successful change. A new backbone conformation is produced by rotating the shorter end of the model protein a random amount in a random direction about a randomly chosen bead. We should note that our previous work with the Dill model using a similar algorithm demonstrated sampling problems for chains longer than 48 beads. New solvent orientations are derived by selecting an individual solvent molecule at random and attempting to randomly change its orientation. This second, solvent "flip" move, is also applied to solvents displaced during a backbone rotation move. Annealing simulations that mimic denaturation are performed as described in previous work (O'Toole and Panagiotopoulos, 1992)

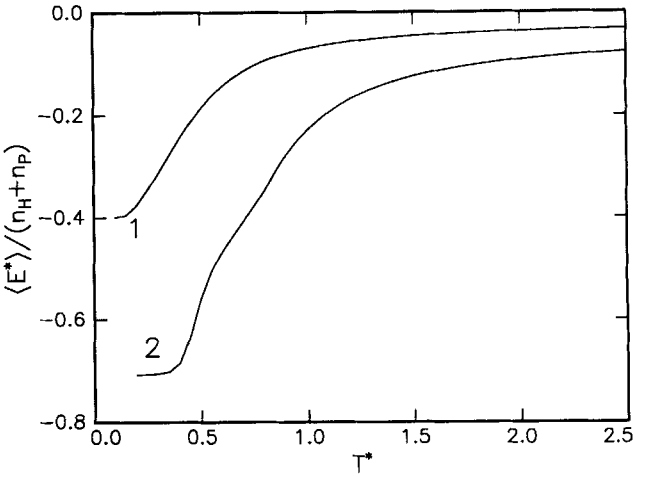

Figure 3. Thermal transition for two sequences in vacuum, calculated using the modified Rosenbluth and Rosenbluth technique.

$1-P_2H_2P_4H_4P_3$; $2-(H_3P_3)_8$ ensemble averaged reduced energy per residue ($\langle E^* \rangle / (n_H + n_P) = \langle E \rangle / (n_H + n_P)\epsilon_{HH}$) is plotted against reduced temperature ($T^* = kT/\epsilon_{HH}$); ϵ_{HH} is the directed hydrophobic-hydrophobic interaction; n_H is the number of hydrophobic beads; n_P is the number of hydrophilic beads.

Results and Discussion

Three sequences were studied: a short 15-bead sequence ($P_2H_2P_4H_4P_3$) and two 48-bead sequences—$(H_3P_3)_8$ and $P_2(H_4P_4)_5H_4P_2$. The 15-bead sequence was constructed randomly, and is 40% hydrophobic. Both 48-bead sequences have been studied previously using the simple Dill model (O'Toole and Panagiotopoulos, 1992). All three sequences were studied in vacuum; the fifteen bead sequence was also studied in a solvent bath. As with our previous work on the Dill model, the modified Rosenbluth and Rosenbluth technique is found to provide efficient sampling of model protein conformations in vacuum at all temperatures. Equilibrium thermal denaturation transitions for two of the three sequences in vacuum—$P_2H_2P_4H_4P_3$ and $(H_3P_3)_8$—calculated using this technique are plotted in Figure 3. These calculations were performed on a CM-5 parallel supercomputer at NCSA in Illinois. The statistical uncertainty for the transition curve calculated with the modified Rosenbluth and Rosenbluth algorithm is of the magnitude of twice the line thickness. The uncertainty was determined from independent runs at identical conditions. The transition for sequence $(H_3P_3)_8$ is the sharpest we have calculated thus far with the directional model, and is significantly sharper than the transition calculated with this sequence for the Dill model (O'Toole and Panagiotopoulos, 1992). However, it is not as sharp as the sharpest transition calculated with the Dill model. There is a "bump" in the transition at approximately $T^* = 0.8$. This "bump" is independent of sample size, within the error of our calculations, and separates the denaturation transition from the single low-energy state found for this sequence into two stages. Two-stage denaturation of this sort is unique within our experience of both this directional model and the simple Dill model (O'Toole and Panagiotopoulos, 1992, 1993), and may disappear with a much larger sample of generated conformations. However, multistage thermal denaturation is well-known experimentally (Privalov, 1982) and indicates that some sections of the protein are more thermally stable than others.

The thermal denaturation transition for the sequence $P_2(H_4P_4)_5H_4P_2$ has also been calculated in vacuum. It is less sharp than the transition for $(H_3P_3)_8$ shown in Figure 3. The transition for the sequence $P_2(H_4P_4)_5H_4P_2$ of the Dill model (O'Toole and Panagiotopoulos, 1992) is significantly sharper than the transition calculated with this sequence using the directional model. Both 48-bead sequences were found to have just one low-energy state in our simulations. The single low-energy state found for the $(H_3P_3)_8$ sequence is shown in Figure 4. This compact conformation has an energy of -34. A reasonable amount of ordering is seen in the structure, with most hydrophobic beads gathering at the core of the structure. The low-energy state found for the sequence $P_2(H_4P_4)_5H_4P_2$ has an energy of -33 and does not exhibit the same amount of ordering as the structure of Figure 4.

The simple solvent "flip" algorithm allows calculation of accurate average energies for solvent baths without model protein backbone. The change in average bath energy with temperature is much more gradual than any of the thermal transitions we have calculated for model proteins in vacuum. The presence of model protein has no detectable effect on the average bath energy. However, the solvent bath appears to have a significant effect on the thermal denaturation transition for the short sequence $P_2H_2P_4H_4P_3$. The transition calculated in a solvent bath of 40^3 solvents for this short sequence is noticeably *less sharp* than the corresponding vacuum transition. Sampling inefficiencies experienced with the Metropolis algorithm described here for significantly longer model proteins in solvent prevent us from making any statement about the general effect of solvent on model protein denaturation transitions. There are reasons why we might expect a decrease in abruptness of thermal denaturation transi-

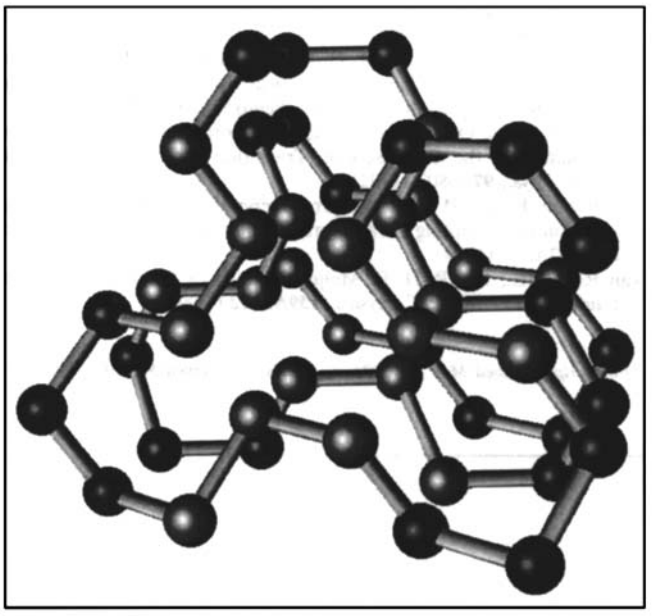

Figure 4. Single low-energy structure found for the sequence $(H_3P_3)_8$.

Arms extending from the beads have been omitted for clarity.

tion for most sequences in solvent baths. It is possible that the interaction set defined above does not represent a proper balance between the interaction of model protein beads with themselves and with solvent. While the model hydrophobic beads interact with the model backbone differently to solvent, the hydrophilic beads do not. In vacuum, a disruption in backbone conformation may result in a significant change in interaction of both hydrophilic and hydrophobic beads with their environment. A similar configurational change in solvent might only result in significant energy change for the hydrophobic beads. Since it is possible for a model hydrophobic bead to be completely exposed to solvent with no resulting repulsion, the net hydrophobic-solvent interaction may be equivalent to the hydrophobic-vacuum interaction. Thus, the energy of the model backbone in solvent may be much less dependent on conformation than is the case in vacuum. More work needs to be done to demonstrate that the model as currently defined will in general result in less abrupt denaturation transitions in solvent, and to investigate the possible explanation offered for such behavior.

Conclusions

In this article we have outlined a directional lattice protein model which includes an orientable solvent, and a vacuum variation where the solvent is removed. We have presented some preliminary results obtained for three sequences: a short random 15 bead sequence, and two 48-bead sequences studied previously with the simple Dill model. Thermal denaturation transitions have been calculated for all three sequences in vacuum, and also in solvent for the short 15-bead sequence. The modified Rosenbluth and Rosenbluth technique is found to provide efficient sampling in vacuum for all sequence lengths described here. The Metropolis algorithm used provides effective sampling for the short sequence. Sharp thermal denaturation transitions are possible in vacuum, although none of the sequences examined thus far displays a transition as sharp as the most abrupt transition calculated with the Dill model. Single low-energy conformations were recovered for both 48-bead sequences in vacuum. A simple solvent "flip" algorithm does provide efficient sampling for the solvent bath, which shows a more gradual reduction in the calculated average energy with increasing temperature than does either sequence in vacuum. The presence of an orientable solvent makes the transition calculated for the short sequence less sharp when compared to the transition this sequence displays in vacuum. Sampling inefficiencies prevent us from drawing any conclusions about the generality of this effect. It is possible that with the interaction set defined for the model may, in general, result in model protein energies in solvent that are less dependent on conformation. More work is needed to investigate this. However, the work described in this article shows that the model strategy outlined here provides a simple means of testing hypothetical mechanisms of protein stabilization in solution.

Acknowledgments

The authors would like to thank Professor Debenedetti for helpful discussions and for providing preprints of papers prior to publication, and are grateful to Professor Dill for providing several preprints. The authors acknowledge the financial support of the Cornell Biotechnology Center and the National Science Foundation through a PYI award to AZP. A generous allocation of time on a CM-5 parallel supercomputer was provided by the National Center for Supercomputing Applications.

Literature Cited

Bell, G. M., "Statistical Mechanics of Water: Lattice Model with Directed Bonding," *J. Phys. C.*, **5**, 889 (1972).

Borick, S. S., and P. G. Debenedetti, "Equilibrium, Stability, and Density Anomalies in a Lattice Model with Core-Softening and Directional Bonding," *J. Phsy. Chem.* **97**, 6292 (1993).

Cantor, C. R., and P. R. Schimmel, *Biophysical Chemistry: Part I. The Conformation of Biological Macromolecules; III. The Behavior of Biological Macromolecules*, W. H. Freeman and Co., New York (1980).

Chan, H. S., and K. A. Dill, "Compact Polymers," *Macromolec.*, **22**, 4559 (1989).

Chan, H. S., and K. A. Dill, "The Effects of Internal Constraints on the Configurations of Chain Molecules," *J. Chem. Phys.* **92**, 3118 (1990).

Chan, H. S., and K. A. Dill, "Sequence Space Soup of Proteins and Copolymers," *J. Chem. Phys.* **95**, 3775 (1991).

Chan, H. S., and K. A. Dill, "The Protein Folding Problem," *Physics Today*, **46**(2), 24 (1993).

Gutin, A. M., and E. I. Shakhnovich, "Ground State of Random Copolymers and the Discrete Random Energy Model," *J. Chem. Phys.*, **98**, 8174 (1993).

King, J. A., "Deciphering the Rules of Protein Folding," *Chem. Eng. News.*, **67**(15), 32 (1989).

Lau, K. F., and K. A. Dill, "A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins," *Macromolec.*, **22**, 3986 (1989).

Meijer, P. H. E., R. Kikuchi, and P. Papon, "Phase Diagram of Water Based on a Lattice Model," *Physica*, **109A**, 365 (1981).

Meijer, P. H. E., R. Kikuchi, and E. Van Royen, "Extended Lattice Model of Water; Two-Sublattice, Two-Orientation Model," *Physica*, **115A**, 124 (1982).

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**, 1087 (1953).

Miller, R., C. A. Danko, M. J. Fasolka, A. C. Balazs, H. S. Chan, and K. A. Dill, "Folding Kinetics of Proteins and Copolymers," *J. Chem. Phys.*, **96**, 768 (1992).

O'Toole, E. M., and A. Z. Panagiotopoulos, "Monte Carlo Simulation of Folding Transitions of Simple Model Proteins Using a Chain Growth Algorithm," *J. Chem. Phys.*, **97**, 8644 (1992).

O'Toole, E. M., and A. Z. Panagiotopoulos, "Effect of Sequence and Intermolecular Interactions on the Number and Nature of Low-Energy States for Simple Model Proteins," *J. Chem. Phys.*, **98**, 3185 (1993).

Privalov, P. L., "Stability of Proteins: Proteins Which Do Not Present a Single Cooperative System," *Adv. Prot. Chem.*, **35**, 1 (1982).

Rosenbluth, M. N., and A. W. Rosenbluth, "Monte Carlo Calculation of the Average Extension of Molecular Chains," *J. Chem. Phys.*, **23**, 356 (1955).

Shakhnovich, E., G. Farztdinov, A. M. Gutin, and M. Karplus, "Protein Folding Bottlenecks: A Lattice Monte Carlo Simulation," *Phys. Rev. Lett.* **67**, 1665 (1991).

Shakhnovich, E. I., and A. M. Gutin, "Formation of Unique Structure in Polypeptide Chains: Theoretical Investigation with the Aid of a Replica Approach," *Biophys. Chemistry*, **34**, 187 (1989).

Szleifer, I., E. M. O'Toole, and A. Z. Panagiotopoulos, "Monte Carlo Simulation of the Collapse-Coil Transition in Homopolymers," *J. Chem. Phys.*, **97**, 6802 (1992).

van Royen, E. and P. H. E. Meijer, "Dependence of the Phase Diagram on the Coupling Parameters in Water-Lattice Models," *Physica*, **127A**, 87 (1984).

van Royen, E., and P. H. E. Meijer, "Limits of Stability of the Extended Water Model," *Physica*, **139A**, 412 (1986).

Manuscript received May 12, 1994, and revision received Sept. 14, 1994.