Thermodynamics of an All-Atom Off-Lattice Model of the Fragment B of *Staphylococcal* Protein A: Implication for the Origin of the Cooperativity of Protein Folding

Yaoqi Zhou* and Apichart Linhananta

Department of Physiology & Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, New York 14214

Received: October 16, 2001

An off-lattice all-atom (except nonpolar hydrogen) model of proteins based on Gō-type discontinuous interactions is developed and applied to study the folding thermodynamics of fragment B of *Staphylococcal* protein A. Unlike simpler C_α based off-lattice models, which fold into a molten globule-like state from the coil state, the new model transits directly to the native state. The transition is strongly first-order-like and the dynamics of the native state is in approximate agreement with experimental data. The results suggest that a large well-packed *solid* core is the origin of the folding cooperativity of proteins.

Introduction

It has been well established that the folding transition of many proteins, in particular, of small globule proteins is first order-like (it is a two-state transition with no detectable intermediates).[1] Proteins can fold cooperatively either from a coil or from a molten globule state with variable secondary structural contents.[2] The origin of cooperativity, however, is not fully understood. The proposed origins of protein's two-state behavior range from helix-coil transitions,[3] heteropolymer collapse,[4,5] side chain packing,[6] to the existence of elementary folding units.[7] Although simplified models can exhibit first-order-like transitions, their interpretations vary. In sophisticated lattice models, the cooperativity arises from multibody interactions,[8,9] whereas different mechanisms (collective orientational rearrangement versus cooperative native-contact formation) are suggested for lattice models with and without side chains.[10] Studies of C_α based (without side chains) off-lattice models, on the other hand, failed to produce a first order coil-to-native folding transition even for highly optimized sequences.[11−13] Instead, a strong transition to a molten globule state, followed by a weak folding transition, is observed.[11−15] To better understand folding thermodynamics, as well as the kinetics, clearly, there is a need for a more accurate off-lattice model.

Atomic level models of proteins in explicit solvents, based on empirical force fields (such as CHARMM[16]), have provided useful information on short-time dynamics[17] and free-energy surfaces.[18] However, unlike simple reduced models, detailed thermodynamic studies of proteins, are not yet feasible. This paper introduces an off-lattice Gō-type all-atom (except nonpolar hydrogens) model with discontinuous interactions. The model is an extension of the simpler C_α-based model developed by Zhou and Karplus to study protein folding.[11−13] Off-lattice models employing discontinuous interactions have the advantage of allowing more realistic dynamics than lattice models without requiring inordinate amount of computation time. We shall illustrate that the all-atom off-lattice model of a three-helix bundle protein accurately describes the dynamical properties of the native state. The inclusion of the side chains ensures a well-packed solidlike core and results in a first order-like folding transition.

Methods

The fragment B of *Staphylococcal* protein A, a 46 residue three-helix bundle protein, is chosen for this study because it is a well-studied experimental[19−21] and theoretical[12−14,18,22−26] model for the folding of helical proteins. The setup for our theoretical model is as follows. The initial heavy-atom positions of the native state of the model protein were obtained from a nuclear magnetic resonance (nmr) study.[27] The NH2-terminus and *N*-methyl COOH-terminus of the polypeptide were capped with acetyl and amine groups, respectively. The total number of atoms is 459. The initial positions of polar hydrogens were generated by the CHARMM program[16] and were minimized for 100 steps to remove bad contacts using the CHARMM polar hydrogen parameter set 19 with distance dependent dielectric constant.[28]

In the present model, all heavy atoms and polar hydrogens (bonded to N or O) are represented by a bead. Two bonded atoms *i* and *j* as well as any 1,3 angle-constrained pair and 1,4 aromatic carbon pair, are constrained to a center-to-center distance between $0.9\sigma_{ij}$ and $1.1\sigma_{ij}$, where σ_{ij} is the atomic separation obtained by the CHARMM energy minimization. This constraint is accomplished by an infinitely deep square-well potential

$$u_{i,j}^{\mathrm{bond}}(r) = \begin{Bmatrix} \infty & r < 0.9\sigma_{ij} \\ 0 & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\ \infty & r < 1.1\sigma_{ij} \end{Bmatrix} \quad (1)$$

We also introduce a "bond" potential for improper dihedral angles to maintain chirality about a tetrahedral extended heavy atom and certain planar atoms. The potential has the form

$$u_{\omega}^{\mathrm{improp}} = \begin{Bmatrix} \infty & \omega > \omega_0 + 20^\circ \\ 0 & \omega_0 - 20^\circ < \omega < \omega_0 + 20^\circ \\ \infty & \omega < \omega_0 - 20^\circ \end{Bmatrix} \quad (2)$$

A large 20° fluctuation was used so that the protein can fold faster in folding kinetic studies (Linhananta and Zhou, in

* To whom correspondence should be addressed. Phone: (716) 829−2985. Fax: (716) 829−2344. Email: yqzhou@buffalo.edu.

10.1021/jp013824r CCC: $22.00
Published on Web 01/19/2002

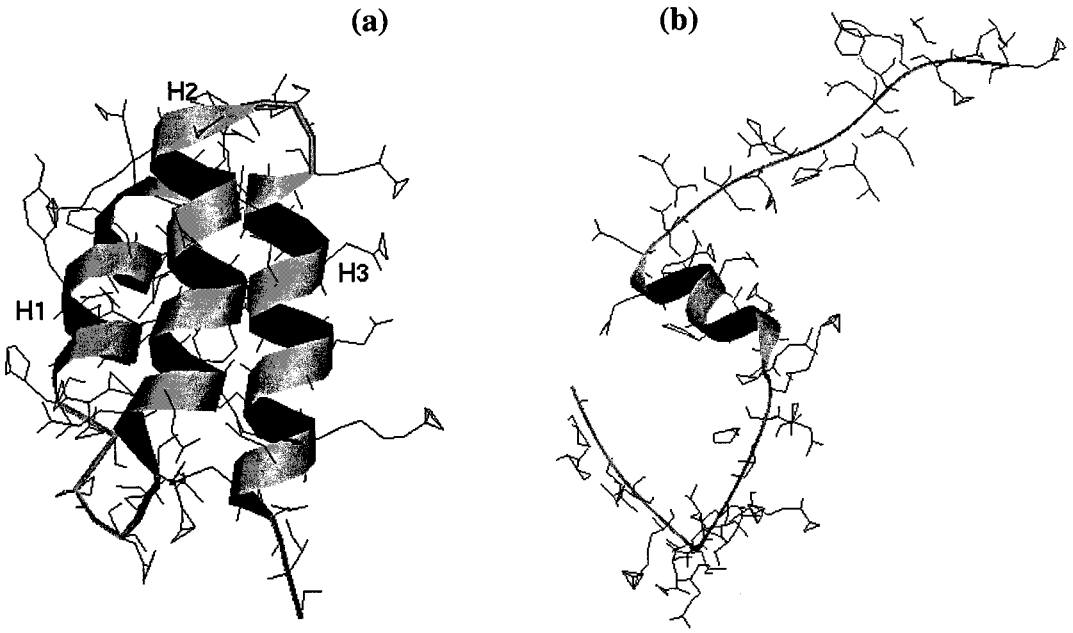

Figure 1. (a) Ribbon representation of the global minimum structure of an all-atom off-lattice model of the fragment B of *Staphylococcal* protein A. The atomic positions are indicated by linked sticks which represent atomic bonds. (b) Instantaneous structure at $T^* = 4.0$. A small helical segment is visible along the coil. These residual helical contents are constantly being formed and broken at high temperature. Drawn using *Molscript*.[47]

preparation). A smaller value will lead to a weaker folding transition but will not affect the overall results presented here. The improper angles ω are identified by using the CHARMM 19 parameter set.[16] In eq 2, $\omega_0 = 35.264\,39°$ for chiral-constrained atoms (such as an α carbon without explicit hydrogens) and $\omega_0 = 0°$ for planar-constrained atoms (such as a carbonyl carbon). The improper dihedral potential u_ω^{improp} preserves the L-form chirality of amino acids and mimics some of the rigidity of peptide units. A nonbonded *i*,*j* pair interacts by a hard-core and square-well potential

$$u_{i,j}(r) = \begin{cases} \infty & r < 0.8\sigma_{ij}^{vdW} \\ B_{ij}\epsilon & 0.8\sigma_{ij}^{vdW} < r < 1.2\sigma_{ij}^{vdW} \\ 0 & r > 1.2\sigma_{ij}^{vdW} \end{cases} \qquad (3)$$

where σ_{ij}^{vdW} are the van der Waals parameters from the CHARMM polar hydrogen parameter set 19[28] and B_{ij} is the yet-to-be-determined interaction strength. The factor of 0.8 for the hard-core diameters is typical for ratio between the diameter of a hard-sphere reference system and the van der Waals parameter found in the Week−Chandler−Andersen perturbation theory,[29] whereas a ratio of 1.5 between the square-well and hard-core diameters is typical for systems of small molecules.[30]

Initial hard-core overlaps in the CHARMM minimized structure are removed by a short discontinuous molecular dynamics (DMD) simulation during which the square-well interactions for the contacts found in the NMR structure are set to −100. A large negative energy is used to ensure that the original native contacts are maintained during the process of removing overlapping contacts. The initial hard-core diameter between any two overlapped atoms is set to the distance between the atoms found in the energy-minimized structure. The hard-core diameters are adjusted at each time step until the corresponding actual diameters are reached and the simulation is continued until all initial overlaps are removed. The resulting structure has a root-mean-squared deviation (rmsd) from the NMR averaged structure of only 0.70 Å and is employed as the global minimum structure for the model (Figure 1a). Helix I (H1) includes residue 2 to 11, Helix II includes residue 16 to 30, and Helix III includes residue 33 to 47. In all, there are 3381 square-well atom−atom overlaps (including both backbone and side chain contacts) in the global minimum structure. Henceforth, these shall be referred to as native contacts.

A Gō interaction[31] is employed to ensure that the energy of the native structure shown in Figure 1a is at the global minimum. In a Gō model, a square-well overlap between two atoms results in an interaction energy of $-\epsilon(B_{ij} = -1)$ for a native contact and 0 ($B_{ij} = 0$) otherwise. We use the Gō model because the native topology was found to play a dominant role in the folding rate and the folding transition state of small fast-folding proteins.[11−13,32−36] Henceforth, the internal energy, E^* ($= E/\epsilon$), and temperature, T^* ($= k_BT/\epsilon$), are scaled in units of ϵ.

The DMD simulations for the isolated model protein are performed in a canonical ensemble. Constant temperature is maintained by collisions with ghost-solvent particles. The algorithm and implementation of DMD are described in earlier publications.[12,37] The initial configurations, in all runs, are set to the global minimum structure. The initial velocities are randomly generated from a Maxwell−Boltzmann distribution. The simulations are performed for a total of at least 500 million and up to 1.4 billion collisions near phase transition. Equilibration periods of between one hundred to two hundred million collisions, depending on temperature, are allowed before data collection. Simulations are conducted at $T^* = 1.0$,1.2, 1.5, 1.7, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, and 5.0. Five independent runs with different initial velocities are performed at each temperature to estimate errors.

Results and Discussion

The DMD simulations show that for reduced temperature $T^* < 3.3$, the protein assumes a low-energy nativelike structure (Figure 1a), whereas for $T^* > 3.6$ the protein rapidly unfolds (in as little as 10 million collisions) to a high-energy coillike state with transient helical segments (Figure 1b). In Figure 2a the reduced heat capacity, C_v^* ($= C_v/k_B$), and internal energy, E^*, are plotted as a function of reduced temperature, T^*. The

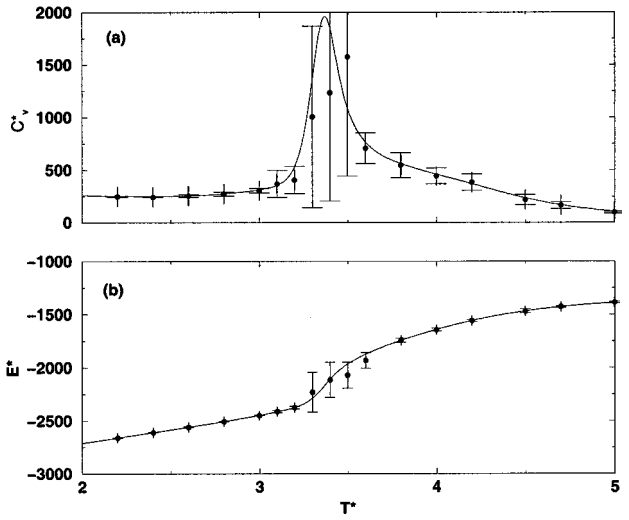

Figure 2. (a) Reduced heat capacity (C_v^*) of the all-atom model as a function of reduced temperature. (b) The reduced energy (E^*) of the all-atom model as a function of the reduced temperature. The strong fluctuation of C_v^* near the transition temperature is due to the coexistence of the coil and folded states.

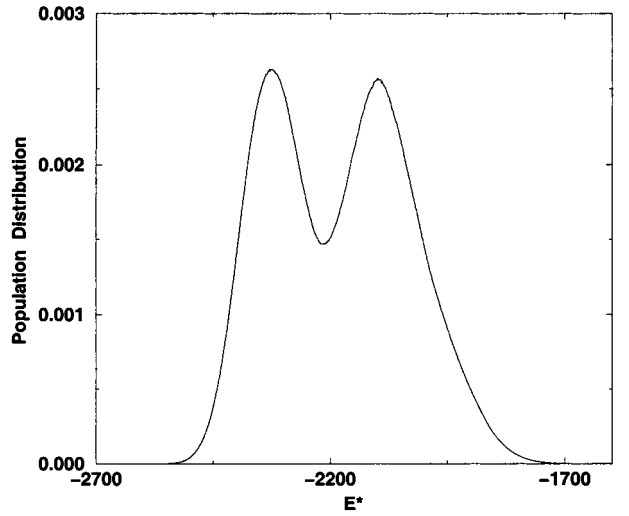

Figure 3. Population distribution at the transition temperature ($T^* = 3.35$) obtained by the weighted-histogram method.[43,37]

sigmoidal increase in E^* in a very narrow T^* range from 3.3 to 3.6, and the sharp peak in C_v^* at $T^* = 3.35$ suggest a strong first order-like transition. Indeed, weighted-histogram analysis at the transition temperature $T^* = 3.35$ shows a bimodal population distribution (Figure 3) indicating a two-state transition. The transition is accompanied by a large increase in the radius of gyration from about 10.5 Å to about 20.0 Å.

It is of interest to compare the C_α model developed by Zhou and Karplus[12,13] with the all-atom model proposed here; both employed discontinuous Gō interactions. The global minimum structure of the C_α model was also optimized to yield a three-helix bundle structure similar to Fragment B of *Staphylococcal* protein A. The global minimum energy for the 46-bead model is -254ϵ. The number of square-well contacts per bead is 5.5, which is much smaller than 7.4 for the all-atom model. This suggests that the all-atom model is a lot more compact than the simple backbone model.

Figure 4 shows the reduced heat capacity and energy as a function of reduced temperature for the C_α three-helix bundle.[12,13] The single transition observed for the all-atom model is replaced by multiple transitions that correspond to coil to molten globule,

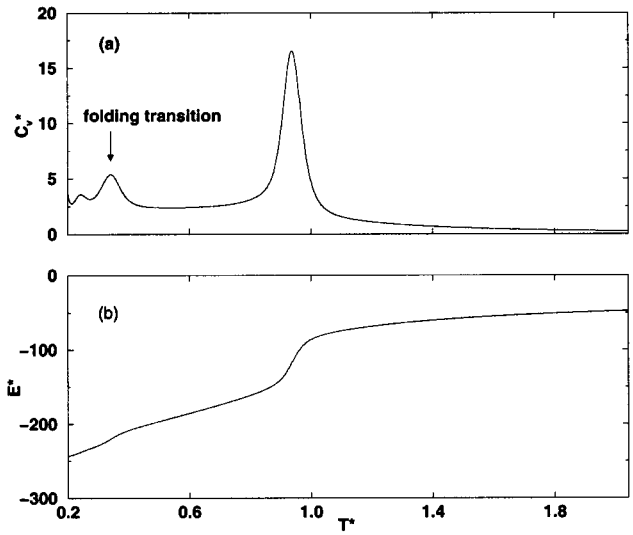

Figure 4. Reduced heat capacity (C_v^*), (a), and the reduced energy (E^*), (b), of the C_α based model[12] as a function of reduced temperature. The results are for a Gō interaction with the bias gap $g = 1$ of Ref. 12. Three C_v peaks from right to left shown in (a) correspond to the collapse transition from coil to molten globule, the folding transition from molten globule to surface-molten solid, and the surface-molten solid to solid transition.

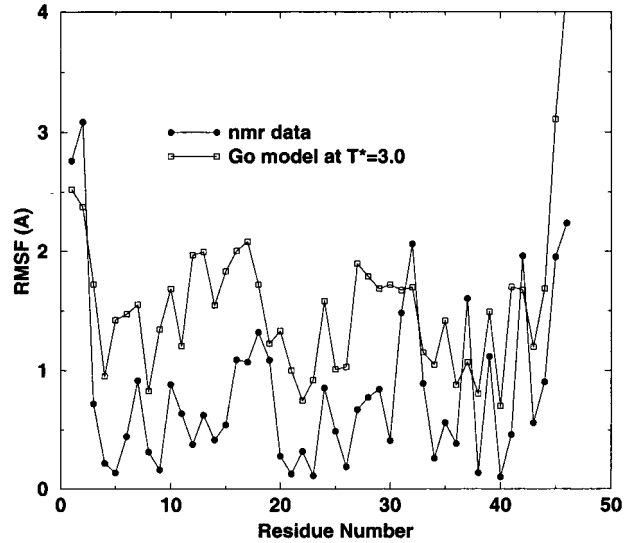

Figure 5. Root-mean-squared fluctuation as function of residue index from the model calculation (open squares) is compared with that from the 10 NMR model structures (filled circles).[27]

molten globule to surface-molten solid (native), and frozen transition.[11−13] Because the global minimum energy of the all-atom model is about a factor of 10 lower than that of the C_α model, a simple scaling suggests that the single transition at $T^* = 3.35$ for the all-atom model corresponds to the folding transition at $T^* = 0.3$ for the 46-bead model. That is, the all-atom model protein folds directly from coil to the native state.

To verify that the protein is in the native state below the transition temperature, structural and dynamical analysis is performed at $T^* = 3.0$. The rmsd from the NMR structure is only 2.5 Å for all heavy atoms and 1.9 Å for the main chain atoms. Similar rmsd values were obtained in an all-atom simulation of fragment B of protein A in explicit water solvent.[25] Figure 5 compares the root-mean-squared fluctuations (rmsf) of the model with the rmsf of the 10 NMR model structures.[27] The correlation coefficient of 0.68 between the two

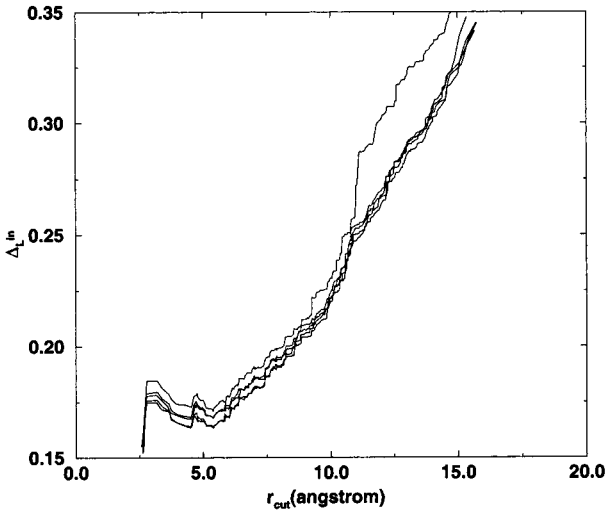

Figure 6. Generalized Lindemann disorder parameter ($\Delta_L^{in}(r_{cut})$) as a function of the distance from the geometrical center, r_{cut} of the protein. Five independent runs at $T^* = 3.0$ are shown in the figure.

sets of data shows that the all-atom model can reproduce, to an acceptable degree of accuracy, the atomic fluctuations of a protein.

The dynamics of the native state can be characterized by the Lindemann parameter, Δ_L.[38] For an inhomogeneous system, we define the generalized Lindemann parameter $\Delta_L^{in}(r_{cut})$[39,40]

$$\Delta_L^{in}(r_{cut}) = \sqrt{\sum_{i,r_i<r_{cut}} \frac{(\Delta_L^i)^2}{N(r_{cut})}}, \Delta_L^i = \frac{\sqrt{\langle \Delta r_i^2 \rangle}}{a'}$$

where r_i and $\langle \Delta r_i^2 \rangle$ are the distance and the rmsf of the i^{th} atom from the geometrical center of the protein, respectively, and a' is the most-probable nonbonded near-neighbor distance. The summation for $\Delta_L^{in}(r_{cut})$ is over all heavy atoms (except those atoms with rotational equivalent positions such as $C_{\gamma 1}$ and $C_{\gamma 2}$ in Valine) within the cutoff distance r_{cut}, with $N(r_{cut})$ being the number of atoms within r_{cut}. The variable a' was chosen to be 4.5 Å as in earlier studies of crambin, myoglobin, and barnase.[40] Figure 6 shows that $\Delta_L^{in}(r_{cut}) \approx 0.17$ for $r_{cut} < 7$ Å at $T^* = 3.0$. Because the typical critical values are 0.13−0.18 depending on the crystal types,[41,42] the core region of the model protein at $T^* = 3.0$ is solidlike. Regions far from the geometrical center have $\Delta_L^{in}(r_{cut})$ ranging from 0.18 to 0.35, indicating large liquidlike fluctuations on the surface. Hence, the single transition observed for the all-atom model is indeed a folding transition between coil and a surface-molten solid state (or, the native state).[11,40]

The strength of the folding transition for the simple all-atom model of the three-helix bundle can be estimated in real units using simple scaling arguments. Scaling the transition temperature for the model ($T^* = 3.35$) to a typical denaturation temperature of ∼350K, gives $\epsilon = 0.875$ kJ/mol. Standard weighted histogram analysis[43,37] found a change in Helmholtz free energy of $137\,\epsilon = 119.9$ kJ/mol from $T^* = 3.3$ to 3.6 which is larger than the measured values of the change in Gibbs free energy of 29 kJ/mol.[19] The higher value is expected since the model protein is energetically optimized (no nonnative interactions) to the native structure of fragment B of *Staphylococcal* protein A. The same scaling method gives a change in energy of $458\,\epsilon = 401$ kJ/mol for the model protein. This is several times larger than the change in the free energy and is characteristic of the large entropic and enthalpic compensation of a folding transition.[1]

The direct coil-to-native transition agrees with experiments on fragment B of *Staphylococcal* protein A.[21] This is different from the phase behavior predicted by the C_α-based models[11,14] where the equilibrium molten globule state is stable over a wide temperature range. It should be emphasized that the existence of the molten globule state and multiple transitions are not unique to models with discontinuous square-well interactions. A weak folding transition and a stable molten globule state were also observed for the C_α models of a three-helix bundle[14] and a four-helix bundle[15] based on Lennard−Jones interactions. Thus, the inclusion of side chains, in this new all-atom model, ensures a well-packed *solid* core resulting in correct thermodynamic behavior for this particular protein.

We propose that the well-packed solidlike core of the native state is the origin of the two-state character of the folding transition. This explains the fact that proteins can cooperatively fold either from a coil or a molten-globule state, analogous to the first-order gas-to-solid and liquid-to-solid transitions in bulk systems. It also provides a link to the first-order-like liquid-to-surface-molten solid transition found in homopolymers[37,44] and isolated clusters.[45] Side chain packing[6] is essential for the folding cooperativity from the molten globule state in which the mainchain structure is more-or-less established. However, for a system lacking a stable molten globule state, side chain and mainchain atoms play an equally important role in cooperativity. Moreover, model proteins without side chains can exhibit a two-state folding transition.[5,11] Heteropolymer collapse[4,5] can be used to interpret cooperativity in folding without molten-globule intermediates. However, such a collapse can also lead to a cooperative or noncooperative transition to a molten-globule state[12] and, thus, is not the fundamental origin for folding cooperativity. Two different mechanisms observed by Hao and Scheraga[10] may be related to the folding to a molten globule and to a native state, respectively.

Not every details of protein thermodynamics are fully accounted for by this model. The C_v curve (Figure 2) shows that the heat capacity of a stable coil state ($T^* \geq 4$) is smaller than that of a stable native state ($T^* \leq 3.1$). This reflects the fact that a random coil has many states with similar energy and, thus, the energy changes slowly with temperature. In real proteins, the denatured state exposes hydrophobic residues to the solvent upon unfolding. This leads to an increased heat capacity at high temperature.[46] Thus, the inclusion of solvation effects is required to achieve a complete description of the thermodynamics of folding. Work in this area as well as on the kinetics of folding is in progress.

Note: After this work was completed, Shimada et al. published a Monte Carlo study using a similar (but simpler) all-heavy atom model.[48]

Acknowledgment. This work was supported by a grant from HHMI to SUNY Buffalo and by the Center for Computational Research and the Keck Center for Computational Biology at SUNY Buffalo.

References and Notes

(1) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167−241.
(2) Ptitsyn, O. B. *Adv. Protein Chem.* **1995**, *47*, 83−230.
(3) Poland, D. C.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers; Statistical Mechanical Theory of Order−Disorder Transitions in Biological Macromolecules*; Academic Press: New York, 1970.
(4) Dill, K. A. *Biochemistry* **1985**, *24*, 1501−1509.
(5) Chan, H. S.; Bromberg, S.; Dill, K. A. *Philos. Trans. R. Soc. London B* **1995**, *348*, 61−70.
(6) Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989**, *28*, 1667−1680.

(7) Murphy, K. P.; Freire, E. *Adv. Protein Chem.* **1992**, *43*, 313−361.
(8) Kolinski, A.; Galazka, W.; Skolnick, J. *Proteins* **1996**, *26*, 271−287.
(9) Kaya, H.; Chan, H. S. *Phys. Rev. Lett.* **2000**, *85*, 4823−4826.
(10) Hao, M. H.; Scheraga, H. A. *J. Mol. Biol.* **1998**, *277*, 973−983.
(11) Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14 429−14 432.
(12) Zhou, Y.; Karplus, M. *J. Mol. Biol.* **1999b**, *293*, 917−951.
(13) Zhou, Y.; Karplus, M. *Nature* **1999a**, *401*, 400−403.
(14) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1490−1494.
(15) Guo, Z. Y.; Brooks, III, C. L. *Biopolymers* **1997**, *42*, 745−757.
(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(17) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−590.
(18) Boczko, E. M.; Brooks, C. L., III *Science* **1995**, *269*, 393−396.
(19) Cedergren, L.; Andersson, R.; Jansson, B.; Uhlen, M.; Nilsson, B. *Prot. Eng.* **1993**, *6*, 441−448.
(20) Bottomley, S. P.; Popplewell, A. G.; Scawen, M.; Wan, T.; Sutton, B. J.; Gore, M. G. *Protein Eng.* **1994**, *7*, 1463−1470.
(21) Bai, Y. W.; Karimi, A.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **1997**, *6*, 1449−1457.
(22) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11 616−11 629.
(23) Guo, Z. Y.; Brooks, III, C. L.; Boczko, E. M. *Proc. Natl. Acad. Sci. (USA)* **1997**, *94*, 10 161−10 166.
(24) Takada, S. *Porteins* **2001**, *42*, 85−98.
(25) Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. (USA)* **2000**, *97*, 133−138.
(26) Kolinski, A.; Skolnick, J. *Proteins.* **1994**, *18*, 353−366.
(27) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665−9672.
(28) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(29) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(30) Sherwood, A. E.; Prausnitz, J. M. *J. Chem. Phys.* **1964**, *41*, 429−437.
(31) Taketomi, H.; Ueda, Y.; Gō, N. *Int. J. Peptide and Protein Res.* **1975**, *7*, 445−459.
(32) Martinez, J. C.; Serrano, L. *Nature Struct. Biol.* **1999**, *6*, 1010−1016.
(33) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nature Struct. Biol.* **1999**, *6*, 1016−1024.
(34) Muñoz, V.; Eaton, W. A. *Proc. Natl Acad. Sci. (USA)* **1999**, *96*, 11 311−11 316.
(35) Takada, S. *Proc. Natl. Acad. Sci. (USA)* **1999**, *96*, 11 698−11 700.
(36) Baker, D. *Nature* **2000**, *405*, 39−42.
(37) Zhou, Y.; Karplus, M.; Wichert, J. M.; Hall, C. K. *J. Chem. Phys.* **1997**, *107*, 10 691−10 708.
(38) Lindemann, F. A. *Physik. Z.* **1910**, *11*, 609−612.
(39) Stillinger, F. H.; Stillinger, D. K. *J. Chem. Phys.* **1990**, *93*, 6013−6024.
(40) Zhou, Y.; Vitkup, D.; Karplus, M. *J. Mol. Biol.* **1999**, *285*, 1371−1375.
(41) Löwen, H. *Phys. Rep.* **1994**, *237*, 249−324.
(42) Stillinger, F. H. *Science* **1995**, *267*, 1935−1939.
(43) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1197.
(44) Zhou, Y.; Hall, C. K.; Karplus, M. *Phys. Rev. Lett.* **1996**, *77*, 2822−2825.
(45) Kunz, R. E.; Berry, R. S. *Phys. Rev. Lett.* **1993**, *71*, 3987.
(46) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1−63.
(47) Kraulis, P. *J. Applied Cryst.* **1991**, *24*, 946−950.
(48) Shimada, J.; Kussell, E. L.; Shakhnovich, E. I. *J. Mol. Biol.* **2000**, *308*, 79−95.