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 Cα 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.

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 experimental19–21 and theoretical12–14,18,22–26 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.23 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 program16 and were minimized for 100 steps to remove bad contacts using the CHARMM polar hydrogen parameter set 19 with distance dependent dielectric constant.25

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 of 0.9σij and 1.1σij, where σij is the atomic separation obtained by the CHARMM energy minimization. This constraint is accomplished by an infinitely deep square-well potential

\[
U_{\text{bond}}^{\text{prop}}(r) = \begin{cases} 
\infty & r < 0.9\sigma_{ij} \\
0 & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\
\infty & r > 1.1\sigma_{ij} 
\end{cases}
\] (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

\[
U_{\text{prop}}^{\text{improp}} = \begin{cases} 
\infty & \omega > \omega_0 + 20^\circ \\
0 & \omega_0 - 20^\circ < \omega < \omega_0 + 20^\circ \\
\infty & \omega < \omega_0 - 20^\circ 
\end{cases}
\] (2)

A large 20° fluctuation was used so that the protein can fold faster in folding kinetic studies (Linhananta and Zhou, in...
preparation). A smaller value will lead to a weaker folding transition but will not affect the overall results presented here. The improper angles $\omega$ are identified by using the CHARMM 19 parameter set.\textsuperscript{16} In eq 2, $\omega_0 = 35.264$ for chiral-constrained atoms (such as an $\alpha$ carbon without explicit hydrogens) and $\omega_0 = 0^\circ$ for planar-constrained atoms (such as a carbonyl carbon). The improper dihedral potential $u_{\text{improp}}$ 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

$$u_{i,j}(r) = \begin{cases} \infty & r < 0.8\sigma_{ij}^{\text{sdw}} \\ B_{ij} \varepsilon_{ij}^{\text{sdw}} & 0.8\sigma_{ij}^{\text{sdw}} < r < 1.2\sigma_{ij}^{\text{sdw}} \\ 0 & r \geq 1.2\sigma_{ij}^{\text{sdw}} \end{cases}$$ \hspace{1cm} (3)$$

where $\sigma_{ij}^{\text{sdw}}$ are the van der Waals parameters from the CHARMM polar hydrogen parameter set 19\textsuperscript{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,\textsuperscript{29} whereas a ratio of 1.5 between the square-well and hard-core diameters is typical for systems of small molecules.\textsuperscript{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\textsuperscript{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 $-\varepsilon (B_{ij} = -1)$ 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.\textsuperscript{11–13,32–36} Henceforth, the internal energy, $E^*(=E/\varepsilon)$, and temperature, $T^* (=k_BT/\varepsilon)$, are scaled in units of $\varepsilon$.

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.\textsuperscript{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, $C_p^*(=C_p/k_B)$, and internal energy, $E^*$, are plotted as a function of reduced temperature, $T^*$. The
sigmoidal increase in $E^*$ in a very narrow $T^*$ range from 3.3 to 3.6, and the sharp peak in $C_V^*/E^*$ 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 $C_R$ model developed by Zhou and Karplus\cite{12,13} with the all-atom model proposed here; both employed discontinuous Go\cite{12} interactions. The global minimum structure of the $C_R$ 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 $-254$. 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 $C_R$ based model\cite{12} as a function of reduced temperature. The results are for a Go 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 compares the root-mean-squared fluctuations (rmsf) of the model with the rmsf of the 10 NMR model structures.\cite{27} The correlation coefficient of 0.68 between the two

molten globule to surface-molten solid (native), and frozen transition.\cite{11,13} Because the global minimum energy of the all-atom model is about a factor of 10 lower than that of the $C_R$ 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.\cite{25} Figure 5 compares the root-mean-squared fluctuations (rmsf) of the model with the rmsf of the 10 NMR model structures.\cite{27} The correlation coefficient of 0.68 between the two

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, $\Delta L$. For an inhomogeneous system, we
define the generalized Lindemann parameter $\Delta_L^{(i)}(r_{\text{cut}})^{39,40}$

$$
\Delta_L^{(i)}(r_{\text{cut}}) = \sqrt{\frac{1}{N} \sum_{i \leq N \left( r_{\text{cut}} \right)} \left( \Delta r_i^2 \right)^2}
$$

where $r_i$ and $\left( \Delta r_i^2 \right)$ are the distance and the rmsf of the $i$th atom
from the geometrical center of the protein, respectively, and $\alpha'$
the most-probable nonbonded near-distance. The summation for $\Delta_L^{(i)}(r_{\text{cut}})$ is over all heavy atoms (except those atoms with rotational equivalent positions such as $C_\text{a}$ and $C_\text{b}$ in Valine) within the cutoff distance $r_{\text{cut}}$, with $N\left( r_{\text{cut}} \right)$ being the number of atoms within $r_{\text{cut}}$. The variable $\alpha'$ was chosen to be
4.5 Å as in earlier studies of crambin, myoglobin, and barnase.40

Figure 6 shows that $\Delta_L^{(i)}(r_{\text{cut}})$ is almost 0.17 for $r_{\text{cut}} < 7$ Å at $T^* = 3.0$.
Because the typical critical values are 0.13–0.18 depending on
the crystal types,31,42 the core region of the model protein at $T^* = 3.0$ is solidlike. Regions far from the geometrical center have
$\Delta_L^{(i)}(r_{\text{cut}})$ 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 temper-
ature for the model ($T^* = 3.35$) to a typical denaturation temperature of $\sim 350\text{K}$, gives $\epsilon = 0.875 \text{ kJ/mol}$. Standard
weighted histogram analysis33,37 found a change in Helmholtz
free energy of 137 $\epsilon = 119.9 \text{ kJ/mol}$ 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 $\epsilon = 401 \text{ kJ/mol}$ 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.3

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 $C_V$-based models31,14
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 $C_V$ models of a three-helix bundle44 and
a four-helix bundle45 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 thermo-
dynamic 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 homopolymers37,44 and
isolated clusters.45 Side chain packing is essential for the folding
cooperativity from the molten globule state in which the main-
chain 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.31,11 Heteropolymer collapse4,5 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 state12 and, thus, is not the fundamental origin for folding
cooperativity. Two different mechanisms observed by Hao and
Scheraga 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

(3) Poland, D. C.; Scheraga, H. A. Theory of Helix-Coil Transitions
Origin of the Cooperativity of Protein Folding
