JCTC Journal of Chemical Theory and Computation

Dynamic Formation and Breaking of Disulfide Bonds in Molecular Dynamics Simulations with the UNRES Force Field

M. Chinchio,[†] C. Czaplewski,[†,‡] A. Liwo,[†,‡] S. Ołdziej,[†,‡] and H. A. Scheraga*[,†]

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, and Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

Received April 6, 2007

Abstract: Many proteins contain disulfide bonds that are usually essential for maintaining function and a stable structure. Several algorithms attempt to predict the arrangement of disulfide bonds in the context of protein structure prediction, but none can simulate the entire process of oxidative folding, including dynamic formation and breaking of disulfide bonds. In this work, a potential function developed to model disulfide bonds is coupled with the united-residue (UNRES) force field, and used in both canonical and replica exchange molecular dynamics simulations to produce complete oxidative folding pathways. The potential function is obtained by introducing a transition barrier that separates the bonded and nonbonded states of the half-cystine residues. Tests on several helical proteins show that improved predictions are obtained when dynamic disulfide-bond formation and breaking are considered. The effect of the disulfide bonds on the folding kinetics is also investigated, particularly their role in stabilizing folding intermediates, resulting in slower folding.

1. Introduction

Disulfide bonds are often present in the native conformations of proteins and contribute to their stability and function. In some cases the disulfide bonds are essential for maintaining the structure of the protein,[1] while for other proteins some bonds can be broken without causing drastic conformational changes.[2−4] Oxidative folding is a very complex process, by which a fully reduced and unfolded protein reaches its native structure (conformational folding) with all native disulfide bonds. Over the years, experimentally determined folding pathways for different proteins have shown a high degree of diversity in the number and type of intermediate states that appear during folding.[1,5−8] Two limiting-case mechanisms have been described:[9] the folded-precursor mechanism, in which conformational folding precedes the formation of the native disulfide bonds, and the quasi-stochastic mechanism, in which disulfide bonds (native and non-native) form in unfolded conformations. Evidence has been found for both,[9,10] and for many proteins the correct mechanism probably lies somewhere in between.

Early theoretical work identified the decreased entropy of the unfolded state as the main source of the increased stability observed for disulfide-bonded proteins.[11−14] In this chain-entropy model, the effect becomes stronger as the sequence separation of the pair of linked residues grows. However, individual proteins can deviate significantly from this ideal behavior, as evidenced by the difficulties encountered in engineering disulfide-stabilized proteins.[15−19] In many cases, one cannot ignore the effect of the introduction of a disulfide bond on the folded state, particularly when the protein structure is perturbed and strained by this bond.

Various aspects of this problem have been attacked computationally using both Monte Carlo and molecular dynamics techniques, and the methods employed have ranged from very simplified lattice Gō-like models to the most

* Corresponding author phone: (607)255-4034; fax: (607)254-4700; e-mail: has5@cornell.edu.
[†] Cornell University.
[‡] University of Gdańsk.

10.1021/ct7000842 CCC: $37.00
Published on Web 06/19/2007

detailed all-atom representations.[17,20−23] However, in most cases the disulfide bonds are not allowed to form and break during the simulations, but rather they are set at the beginning and remain fixed throughout. A few approaches have attempted to model this process dynamically, either by employing a highly simplified 2-D lattice representation[24] or by restricting the dynamics to the packing of preassigned secondary structure elements.[25] While several disulfide-bond prediction algorithms exist,[26−30] the first general approach to protein structure prediction including dynamic disulfide-bond formation and breaking was recently proposed by some of the present authors.[31] In that approach, no assumptions were made as to the positions of native disulfides, and predictions were produced by an unbiased global search for an energy minimum, based on local energy minimization. The energy function and united-residue (UNRES) model employed were previous versions of those presented here, which will be described in more detail in later sections. However, due to its use of local energy minimization in the global search procedure, thermodynamic or kinetic information could not be obtained directly. The work presented here addresses this limitation by using the recently developed molecular dynamics algorithm applied to the UNRES model[32] and introducing dynamic formation and breaking of disulfide bonds. The method is tested on four α-helical proteins, covering a range of fold types and various arrangements of disulfide bonds. The results show that the addition of disulfide bonds can significantly improve the quality of blind structure predictions while, at the same time, providing insight into the role of disulfide bonds in the stabilization of proteins and their effect on the kinetics of folding.

2. Methods

2.1. The UNRES Force Field. In the UNRES model,[33−47] a polypeptide chain is represented by a sequence of α-carbon (C^α) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α-carbons. Only these united peptide groups and the united side chains serve as interaction sites, the α-carbons serving only to define the chain geometry, as shown in Figure 1.

The UNRES force field has been derived as a Restricted Free Energy (RFE) function of an all-atom polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system (i.e., the degrees of freedom of the solvent, the dihedral angles χ for rotation about the bonds in the side chains, and the torsional angles λ for rotation of the peptide groups about the $C^\alpha\cdots C^\alpha$ virtual bonds).[37,38,48] The RFE is further decomposed into factors arising from interactions within and between a given number of united interaction sites.[38] Expansion of the factors into generalized Kubo cumulants[49] facilitated the derivation of approximate analytical expressions for the respective terms,[37,38] including the *multibody* or *correlation* terms. The theoretical basis of the force field is described in detail elsewhere.[38]

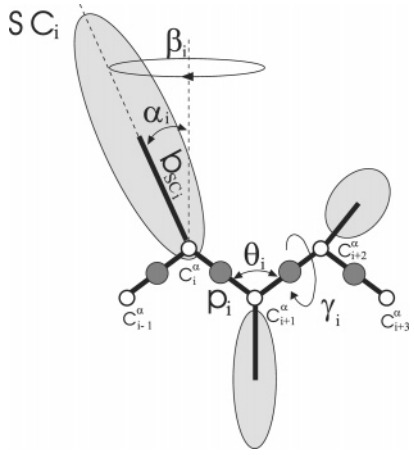

Figure 1. The UNRES model of polypeptide chains. The interaction sites are side-chain centroids of different sizes (SC) and peptide-bond centers (*p*) indicated by shaded circles, whereas the α-carbon atoms (small empty circles) are introduced only to assist in defining the geometry. The virtual $C^\alpha\cdots C^\alpha$ bonds have a variable length centered around 3.8 Å, corresponding to a planar trans peptide group; the virtual-bond (θ) and dihedral (γ) angles are variable. Each side chain is attached to the corresponding α-carbon with a variable "bond length", b_{SC_i}, variable "bond angle", α_{SC_i}, formed by SC_i and the bisector of the angle defined by C^α_{i-1}, C^α_i, and C^α_{i+1}, and with a variable "dihedral angle" β_{SC_i} of counterclockwise rotation about the bisector, starting from the right side of the C^α_{i-1}, C^α_i, C^α_{i+1} frame.

The energy of the virtual-bond chain is expressed by eq 1:

$$U = w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_i p_j} + w_{pp}^{VDW}\sum_{i<j-1} U_{p_i p_j}^{VDW} + w_{pp}^{el} f_2(T)\sum_{i<j-1} U_{p_i p_j}^{el} + w_{tor} f_2(T)\sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T)\sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + \sum_{m=3,4} w_{corr}^{(m)} f_m(T) U_{corr}^{(m)} + \sum_{m=3,4} w_{turn}^{(m)} f_m(T) U_{turn}^{(m)} + w_{bond}\sum_{i=1}^{nbond} U_{bond}(d_i) \quad (1)$$

The term $U_{SC_iSC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{SC_i p_j}$ denotes the excluded-volume potential of the side-chain−peptide-group interactions. The peptide-group interaction potential is split into two parts: the Lennard-Jones interaction energy between peptide group centers ($U_{p_i p_j}^{VDW}$) and the average electrostatic energy between peptide group dipoles ($U_{p_i p_j}^{el}$); the second of these terms accounts for the tendency to form backbone hydrogen bonds between peptide groups p_i and p_j. U_{tor}, U_{tord}, U_b, and U_{rot} are the virtual-bond torsional terms, virtual-bond double-torsional terms, virtual-bond angle bending terms, and side-chain rotamer terms; these terms account for the local propensities of the polypeptide chain. The terms $U_{corr}^{(m)}$ represent the *correlation* or *multibody*

contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the terms $U_{\text{turn}}^{(m)}$ are correlation contributions involving m consecutive peptide groups; they are, therefore, termed turn contributions. The correlation contributions were derived[37,38] from a generalized-cumulant expansion[49] of the restricted free energy (RFE) of the system consisting of the polypeptide chain and the surrounding solvent. The multibody terms are indispensable for reproduction of regular α-helical and β-sheet structures. The terms $U_{\text{bond}}(d_i)$, where d_i is the length of the ith virtual bond, are simple harmonic potentials of virtual bond distortions, introduced for the molecular dynamics implementation, and nbond is the number of virtual bonds. The $f_m(T)$ terms express the temperature dependence of the restricted free energy function. Their main effect is to reduce the weight of multibody terms at high temperatures, thus preventing the premature formation of secondary structure, characteristic of older versions of this force field.

The internal parameters of $U_{p_i\,p_j}$, U_{tor}, U_{tord}, $U_{\text{corr}}^{(m)}$, and $U_{\text{turn}}^{(m)}$ were derived by fitting the analytical expressions to the RFE surfaces of model systems computed by quantum mechanics at the MP2/6-31G** ab initio level,[42,43] while the parameters of $U_{\text{SC}_i\text{SC}_j}$, $U_{\text{SC}_i\,p_j}$, U_b, and U_{rot} were derived by fitting the calculated distribution functions to those determined from the PDB;[36] work is currently in progress to obtain these parameters from quantum mechanical ab initio calculations of the potentials of mean force of appropriate model systems. The w's (the weights of the energy terms), the internal parameters within each cumulant term, and the mean free energies of side-chain interactions of the $U_{\text{SC}_i\text{SC}_j}$ energy term were obtained by a hierarchical optimization[50] of the energy function based on protein 1GAB, a three-helix bundle. This force field is the first version of UNRES parametrized specifically for canonical simulations. It was designed so that folding occurs at physiological temperatures (≈300 K).

In this version of the force field, the side-chain pairwise interaction potential is represented by the orientation dependent Gay−Berne[51] functional form, given by

$$U_{\text{SC}_i\text{SC}_j}^{(\text{GB})} = 4(|\epsilon_{ij}^{(\text{GB})}|x_{ij}^{\,12} - \epsilon_{ij}^{(\text{GB})}x_{ij}^{\,6})$$

$$\epsilon_{ij}^{(\text{GB})} = \epsilon_{ij}\epsilon_{0_{ij}}^{(\text{GB})}$$

$$x_{ij} = \frac{\sigma_{0_{ij}}}{r_{ij} - (\sigma_{ij} - \sigma_{0_{ij}})} \qquad (2)$$

where $\epsilon_{0_{ij}}^{(\text{GB})}$ and $\sigma_{0_{ij}}$ are constant internal parameters ($\epsilon_{0_{ij}}^{(\text{GB})} > 0$ only for hydrophobic−hydrophobic interactions), ϵ_{ij} and σ_{ij} depend on the relative orientation of the two side chains (more details are given in an earlier publication[35]), and r_{ij} is the distance between the side-chain centroids, as shown in Figure 2. For any given orientation, the van der Waals well depth is given by $\epsilon_{ij}^{(\text{GB})}$. However, ϵ_{ij} remains in the range 1.0−2.0 for different orientations; therefore, the well depth is always at least $\epsilon_{0_{ij}}^{(\text{GB})}$.

2.2. Disulfide Bonds in UNRES. Disulfide-bond potentials were first introduced into the UNRES force field as simple harmonic potentials which depended only on the distance between the side-chain centroids of the relevant half-cystines.[31] A more sophisticated model was later developed which takes into account the relative orientation of the two side chains as well as the distance between them.[52] In this model, the disulfide-bond energy is expressed by

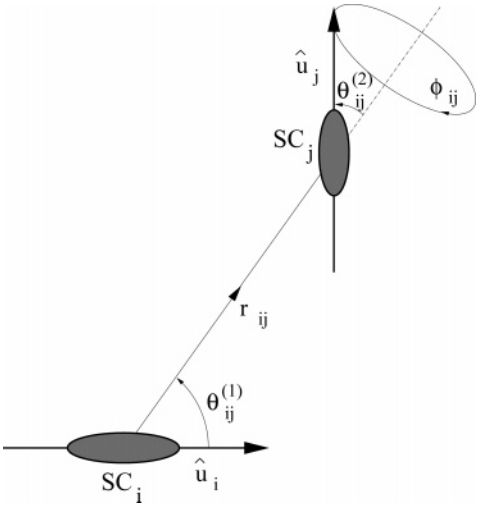

Figure 2. Definition of the orientation of two anisotropic side chains, SC_i and SC_j, represented by ellipsoids of revolution.[35] The relative position of the centers of the side chains are given by the vector $\mathbf{r}_{ij}$ (of length r_{ij}). The principal axes of the ellipsoids are assumed to be colinear with the C^{α}−SC lines; their directions are given by the unit vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$. The variables defining the relative orientations of the ellipsoids are the angles $\theta_{ij}^{(1)}$ (the planar angle between $\hat{\mathbf{u}}_i$ and $\mathbf{r}_{ij}$), $\theta_{ij}^{(2)}$ (the planar angle between $\hat{\mathbf{u}}_j$ and $\mathbf{r}_{ij}$), and ϕ_{ij} (the angle of counterclockwise rotation of the vector $\hat{\mathbf{u}}_j$ about the vector $\mathbf{r}_{ij}$ from the plane defined by $\hat{\mathbf{u}}_i$ and $\mathbf{r}_{ij}$) when looking from the center of SC_j toward the center of SC_i.

$$U_{\text{SC}_i\text{SC}_j}^{(\text{SS})} = \epsilon_0^{(\text{SS})} + \sum_{n=1}^{3} \upsilon_n \zeta_{ij}^n + k_0((\eta_{ij}^{(1)})^2 + (\eta_{ij}^{(2)})^2) + k_1(\eta_{ij}^{(1)} + \eta_{ij}^{(2)})(r_{ij} - d_0) + k_2(r_{ij} - d_0)^2$$

$$\eta_{ij}^{(1)} = 1 - \hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}}_{ij}$$

$$\eta_{ij}^{(2)} = 1 + \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}}_{ij}$$

$$\zeta_{ij} = \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j - (\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}}_{ij})(\hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}}_{ij}) \qquad (3)$$

where the vectors $\hat{\mathbf{u}}$ and $\mathbf{r}$ are defined in Figure 2, $\epsilon_0^{(\text{SS})}$ is an adjustable parameter which defines the well depth, and the internal parameters d_0, υ_n, and k_n were derived based on ab initio calculations of diethyl disulfide at the RHF/6-31G** level.[52] The numerical values of these parameters are summarized in Table 1.

Until now,[31] $U_{\text{SC}_i\text{SC}_j}^{(\text{GB})}$ had to be *replaced* with $U_{\text{SC}_i\text{SC}_j}^{(\text{SS})}$ to calculate the interaction energy between half-cystines forming a disulfide bond. Formation (or breaking) of a disulfide bond was effectively achieved by modifying the relevant term in the UNRES force field. This approach works very well for minimization-based search methods, where the goal is simply to identify the lowest energy conformations. It was successfully applied[31] to the CSA search procedure,[53−55] a

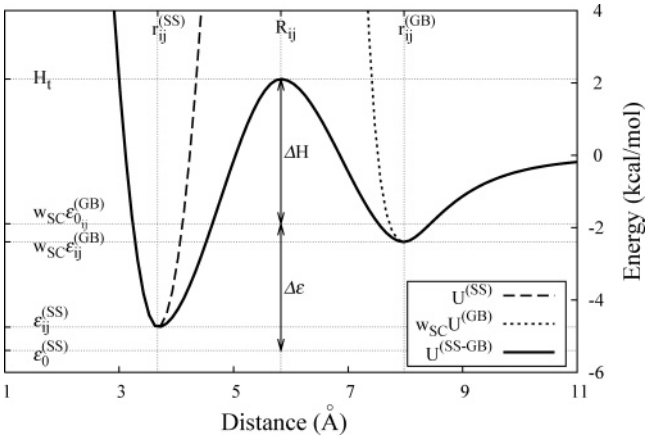

Figure 3. Distance dependence of the energy term $U_{SC_iSC_j}^{(SS-GB)}$, for a representative arrangement of side chains ($\theta_{ij}^{(1)} = 30°$, $\theta_{ij}^{(2)} = 150°$, $\phi_{ij} = 120°$). The dashed and dotted curves represent $U_{SC_iSC_j}^{(SS)}$ and $U_{SC_iSC_j}^{(GB)}$. The most crucial distances (the two minima $r_{ij}^{(SS)}$ and $r_{ij}^{(GB)}$ and the maximum R_{ij}) are shown. The portions of these potentials between $r_{ij}^{(SS)}$ and $r_{ij}^{(GB)}$ are replaced by an artificial transition barrier. The key energy levels are also illustrated, viz., the top of the transition barrier (H_t); the depth ($w_{SC}\epsilon_{ij}^{(GB)}$) of the $U_{SC_iSC_j}^{(GB)}$ potential well and its defining parameter ($\epsilon_{0_{ij}}^{(GB)}$); and the depth of the $U_{SC_iSC_j}^{(SS)}$ potential well ($\epsilon_{ij}^{(SS)}$) and its defining parameter ($\epsilon_0^{(SS)}$). Also shown are the two adjustable parameters ΔH and $\Delta\epsilon$.

Table 1. Internal Parameters of the Disulfide-Bond Potential

parameter	value
d_0	3.78 Å
v_1	−1.08 kcal/mol
v_2	7.61 kcal/mol
v_3	13.70 kcal/mol
k_0	11.0 kcal/mol
k_1	12.0 kcal/mol/Å
k_2	15.1 kcal/mol/Å^2

genetic algorithm which has been used with UNRES for the past several years. CSA was extended to allow for dynamic rearrangement of disulfide bonds during the simulation by introducing new genetic operators to treat the formation and breaking of the bonds.[31] However, in the case of molecular dynamics simulations, this approach cannot be applied easily without destroying the thermodynamic properties of the algorithm. Instead, a new approach is presented here that combines the terms $U_{SC_iSC_j}^{(GB)}$ and $U_{SC_iSC_j}^{(SS)}$ by introducing an artificial transition barrier between the two minima. The resulting potential function, given by eq 4, is shown in Figure 3.

$$U_{SC_iSC_j}^{(SS-GB)} = \begin{cases} U_{SC_iSC_j}^{(SS)} & \text{if } r_{ij} < r_{ij}^{(SS)} \\ \epsilon_{ij}^{(SS)}\, g(r_{ij}; R_{ij}, r_{ij}^{(SS)}) + H_t g(r_{ij}; r_{ij}^{(SS)}, R_{ij}) & \text{if } r_{ij}^{(SS)} \le r_{ij} < R_{ij} \\ H_t g(r_{ij}; r_{ij}^{(GB)}, R_{ij}) + w_{SC}\epsilon_{ij}^{(GB)}\, g(r_{ij}; R_{ij}, r_{ij}^{(GB)}) & \text{if } R_{ij} \le r_{ij} < r_{ij}^{(GB)} \\ w_{SC}U_{SC_iSC_j}^{(GB)} & \text{if } r_{ij} \ge r_{ij}^{(GB)} \end{cases} \quad (4)$$

In the version of UNRES used for the calculations presented here, the values[50] of w_{SC} and $\epsilon_0^{(GB)}$ are 1.35279 and −1.4013, respectively. The side-chain distances corresponding to the minima of the $U_{SC_iSC_j}^{(SS)}$ and $U_{SC_iSC_j}^{(GB)}$ potentials ($r_{ij}^{(SS)}$ and $r_{ij}^{(GB)}$, respectively), and the values of these energy terms at the minima ($\epsilon_{ij}^{(SS)}$ and $w_{SC}\epsilon_{ij}^{(GB)}$) are easily obtained by differentiation of the respective defining functions, eqs 2 and 3. The distance R_{ij} corresponding to the maximum of the transition barrier is chosen arbitrarily to be the midpoint between the two minima.

The potential function $U_{SC_iSC_j}^{(SS-GB)}$ contains only two adjustable parameters, $\Delta\epsilon$ and ΔH, which control the depth of the disulfide-bond potential well and the height of the barrier, respectively:

$$\epsilon_0^{(SS)} = w_{SC}\epsilon_{0_{ij}}^{(GB)} - \Delta\epsilon \quad (5)$$

$$H_t = w_{SC}\epsilon_{0_{ij}}^{(GB)} + \Delta H \quad (6)$$

The function $g(x;a,b)$ was chosen as a sigmoid function between the limits a and b and must have the following properties to define a smooth barrier:

$$g(x = a; a, b) = 0 \quad (7a)$$

$$g(x = b; a, b) = 1 \quad (7b)$$

$$\frac{\partial g}{\partial x}|_{x=a} = 0 \quad (7c)$$

$$\frac{\partial g}{\partial x}|_{x=b} = 0 \quad (7d)$$

The exact form of the function $g(x; a, b)$ was shown to have no significant effect on the performance of the potential function (data not shown); therefore, the following simple sigmoid form was chosen:

$$g(x; a, b) = \left(\frac{x-a}{b-a}\right)^2\left(3 - 2\frac{x-a}{b-a}\right) \quad (8)$$

To dynamically simulate the formation and breaking of disulfide bonds, the new term $U_{SC_iSC_j}^{(SS-GB)}$ is inserted into eq 1 to replace the corresponding $w_{SC}U_{SC_iSC_j}$, for all possible pairs of cysteines. Formation or breaking of a disulfide bond simply corresponds to crossing the barrier at $r_{ij} = R_{ij}$, and occurs naturally during the simulation. The thermodynamic properties of the equilibrium state are expected to be influenced mainly by the parameter $\Delta\epsilon$, which determines the strength of a disulfide bond and will therefore also affect the rate of bond breaking. From our previous work[31] and an analysis of Doig and Williams,[56] we estimate $\Delta\epsilon \approx 3.5$ kcal/mol. The other adjustable parameter (ΔH) determines the rate of disulfide-bond formation and breaking and will therefore strongly influence the kinetics of folding. Both in vivo and in vitro, oxidative folding is controlled by various agents that promote disulfide formation and reshuffling. However, while the folding kinetics can be strongly affected by varying experimental conditions, there is evidence that the main features of the folding mechanism are conserved.[57,58] Since the parameters $\Delta\epsilon$ and ΔH can be thought of as mimicking experimental conditions, this suggests that varying

ΔH within a range that makes the simulations feasible (if it is too large, transitions will never occur) should not change the character of the folding pathways. These effects will be investigated in section 3.

2.3. Simulation Methodology. The kinetics of folding are studied by carrying out molecular dynamics (MD) runs at 300 K. The initial state is obtained by performing a short (100 000 steps) run at 600 K to produce an unfolded state. The simplest method to achieve canonical simulations is to employ the Berendsen thermostat.[59] This approach is not as physically accurate as Langevin dynamics with explicit friction and stochastic forces, but it was shown to produce much faster simulated folding, while still producing qualitatively accurate folding pathways, albeit on a shorter time scale.[60] These features make it ideal for the purpose of testing the new methodology for disulfide-bond formation. As in earlier work,[60] the coupling parameter (τ_T in eq 18 of ref 60) was set to 1 mtu (where 1 mtu = 48.9 fs), and the extension[47] of the velocity Verlet algorithm[61] to include a variable time step was used to integrate the equations of motion, with a basic time step of 0.1 mtu.

In order to enhance sampling and investigate the temperature dependence of various properties of the equilibrium state ensemble (such as free energy and native content), multiplexing replica-exchange molecular dynamics (MREMD[62,63]) runs are also performed. The MREMD method is based on replica-exchange molecular dynamics (REMD[64−68]). In REMD, M canonical MD simulations are performed at different temperatures, covering the range 200−440 K. After every 20 000 steps, an exchange of temperatures is attempted between neighboring trajectories, based on the Metropolis criterion defined by eq 9

$$\Delta = [\beta_{i+1}U(\mathbf{X}_{i+1}, \beta_{i+1}) - \beta_i U(\mathbf{X}_{i+1}, \beta_i)] - [\beta_{i+1}U(\mathbf{X}_i, \beta_{i+1}) - \beta_i U(\mathbf{X}_i, \beta_i)] \quad (9)$$

where $\beta_i = 1/RT_i$, T_i is the temperature of the ith trajectory, $\mathbf{X}_i$ represents all the variables defining the conformation of the ith trajectory (at the time of the exchange), and $U(\mathbf{X}_i, \beta_i)$ is the corresponding UNRES energy. If $\Delta \leq 0$, T_i and T_{i+1} are exchanged; otherwise, the exchange is performed with probability $\exp(-\Delta)$. MREMD builds on REMD by running several trajectories at any given temperature and attempting exchanges between all trajectories at neighboring temperatures. Multiplexing was recently shown to improve the convergence of the simulations significantly.[63]

2.4. Data Analysis. To compute thermodynamic quantities and averages from the results of MREMD simulations, we employ the weighted histogram analysis method (WHAM),[69] a procedure which was recently adapted for use with MREMD simulations and the UNRES potential.[50] Given M simulations at different temperatures producing an ensemble of N conformations, we solve the following set of self-consistent equations for the probabilities (P_i) of all conformations and the dimensionless free energies (f_k) of all simulations

$$\omega_i = -\ln\sum_{k=1}^{M} \exp[-f_k + \beta_k U(\mathbf{X}_i, \beta_k)]$$

$$P_i(\beta_j) = \frac{\exp[-\beta_j U(\mathbf{X}_i, \beta_j)]}{\sum_{k=1}^{M} \exp[-f_k + \beta_k U(\mathbf{X}_i, \beta_k)]} = \exp[\omega_i - \beta_j U(\mathbf{X}_i, \beta_j)]$$

$$f_k = -\ln\sum_{i=1}^{N} P_i(\beta_k) \quad (10)$$

where ω_i can be considered as the entropy of the ith conformation. With these quantities we can compute the average of any quantity A at any temperature T (or $\beta = 1/RT$) over any subset of conformations $\{\mathbf{X}\}$

$$\langle A \rangle_{\beta,\{\mathbf{X}\}} = \frac{1}{Z(\beta,\{\mathbf{X}\})} \sum_{i\in\{\mathbf{X}\}} A_i \exp[\omega_i - \beta U(\mathbf{X}_i, \beta)] \quad (11)$$

where $Z(\beta, \{\mathbf{X}\})$ is the partition function defined by

$$Z(\beta, \{\mathbf{X}\}) = \sum_{i\in\{\mathbf{X}\}} \exp[\omega_i - \beta U(\mathbf{X}_i, \beta)] \quad (12)$$

Based on eq 12, we define the free energy

$$F(\beta, \{\mathbf{X}\}) = -\frac{1}{\beta}\ln Z(\beta, \{\mathbf{X}\}) \quad (13)$$

and the heat capacity

$$C_V(T, \{\mathbf{X}\}) = \frac{\partial}{\partial T}E(T, \{\mathbf{X}\})$$

$$E(T, \{\mathbf{X}\}) = -RT^2\frac{\partial}{\partial T}\ln Z(T, \{\mathbf{X}\}) \quad (14)$$

The heat capacity is particularly useful in identifying the temperature at which the folding transition occurs, which corresponds to a peak in the heat capacity curve when plotted against temperature.

To identify the dominant conformations in the equilibrium state ensemble at any particular temperature, the conformations are clustered with the minimum-variance clustering algorithm,[70,71] using the rms deviation over C^α coordinates (rmsd) as the distance measure. The rmsd cutoff in clustering is adjusted to achieve a balance between the total number of clusters and their compactness. The clusters are then ranked by their probability, which is calculated by adding up the probabilities for the individual conformations making up each cluster, obtained by solving eqs 10. For each cluster, the average rmsd is calculated using eq 11, where A_i is taken as the rmsd of the ith conformation from the native structure.

3. Results

The potential function and methodology described in the previous sections are applied to several proteins, identified by their PDB codes and described in detail in the following sections. The proteins chosen for this work are relatively simple and small but cover different disulfide-bond arrangements and fold types. They were selected primarily to test the various features of the new method, without adding unnecessary complexity. The version of UNRES used here was parametrized on a single small α-helical protein[50]

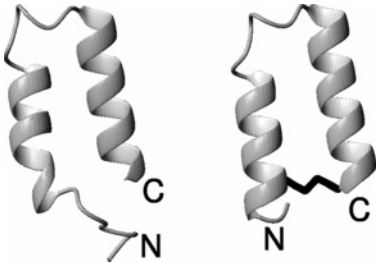

Figure 4. Native conformations[4] of 1ZDB (left) and 1ZDD (right). The disulfide bond in the structure of 1ZDD is highlighted in black. The N- and C-termini are labeled for clarity.

(1GAB) and therefore does not in general perform well on large proteins or proteins containing β structure. As improved versions of the potential are produced (work under way), it will become possible to study more complex systems.

For each protein, several canonical MD and MREMD runs are carried out with different values for the parameters $\Delta\epsilon$ and ΔH. To assess the effect of introducing disulfide bonds, one run (labeled NOSS) is performed with the $U^{(\mathrm{GB})}_{\mathrm{SC}_i\mathrm{SC}_j}$ potential instead of $U^{(\mathrm{SS-GB})}_{\mathrm{SC}_i\mathrm{SC}_j}$, i.e., without allowing for disulfide-bond formation. Each MREMD simulation consists of 100−200 trajectories. The ensemble of conformations used to calculate thermodynamics properties from MREMD simulations is obtained by sampling over the last 8 000 000 steps of each trajectory. The total length of each trajectory is 24 000 000 steps. For each canonical MD simulation, statistics are obtained from 160 trajectories, each 20 000 000 steps long (just under 100 ns).

3.1. Proteins 1ZDB and 1ZDD. Protein 1ZDB is a 38-residue fragment of the B-domain of protein A, comprising its first two helices. Several residues have been mutated to enhance its binding affinity to IgG,[72] and the fragment was shown to fold independently to the same structure as in the full protein.[4] Removal of the first five N-terminal residues further improves binding; we refer to this shorter fragment as 1ZDB*. According to ref 4, 1ZDD is a one-disulfide variant of 1ZDB, showing greatly enhanced thermal stability. The three sequences are

```
1ZDB   :  AVAQSFNMQQQRRFYEALHDPNLNEEQRNAKIKSIRDD

1ZDB*  :      FNMQQQRRFYEALHDPNLNEEQRNAKIKSIRDD

1ZDD   :      FNMQCQRRFYEALHDPNLNEEQRNAKIKSIRDDC
```

The native conformations[4] of proteins 1ZDB and 1ZDD are shown in Figure 4, with the disulfide bond highlighted. The unstructured N-terminal region of 1ZDB, clearly visible in the Figure, is removed to produce 1ZDB*, for which no experimental structure was determined.[4]

The heat capacity curves for 1ZDB, 1ZDB*, and 1ZDD are shown in Figure 5. Despite some differences, all curves point to a well-defined folding transition around 310 K, which justifies the choice of 300 K to study the folded state. The results shown in Table 2, obtained from the MREMD runs, confirm that nativelike conformations are more prevalent for the shorter fragment, consistent with experimental

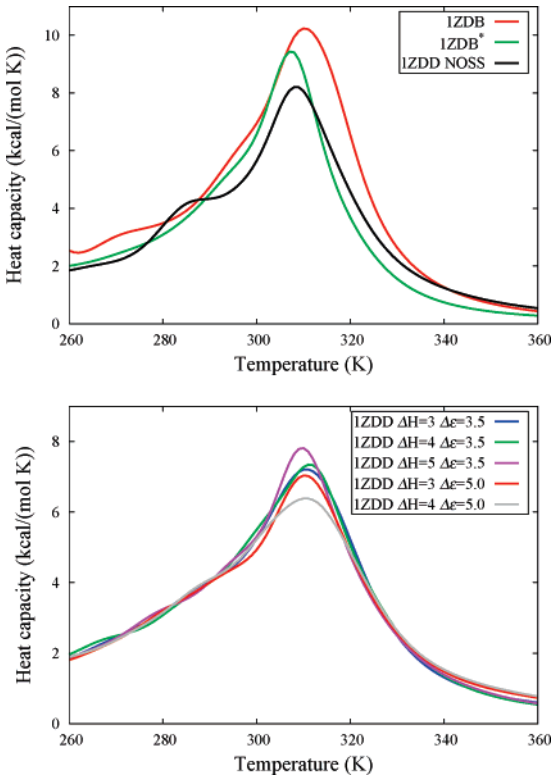

Figure 5. Temperature dependence of the heat capacity for proteins 1ZDB, 1ZDB*, and 1ZDD, with data obtained from MREMD simulations.

Table 2. Percentage of Conformations within a Given RMSD Cutoff from the Corresponding Native Structure in the Equilibrium Ensemble at 300 K[a]

test system	rmsd < 4 Å	rmsd < 5 Å	rmsd < 6 Å
1ZDB	2	16	38
1ZDB*	9	39	63
1ZDD			
NOSS	25	65	89
$\Delta\epsilon = 3.5$			
$\Delta H = 3$	47	78	95
$\Delta H = 4$	49	78	95
$\Delta H = 5$	47	79	95
$\Delta\epsilon = 5.0$			
$\Delta H = 3$	57	84	97
$\Delta H = 4$	52	81	97

[a] Data obtained from MREMD runs.

observations.[72] It is also clear from the results of the NOSS run that greater prevalence of native structures is obtained with the mutation of one residue to cysteine and the addition of another cysteine at the C-terminus, even without considering the formation of the disulfide bond. However, allowing the disulfide bond to form enhances the quality of the prediction significantly, with approximately twice as many conformations found within 4 Å from the native (from 25% in the NOSS run to ≈50% when the disulfide bond is allowed to form).

As discussed in section 2.2, broad thermodynamic properties are affected mainly by the parameter $\Delta\epsilon$ and not by ΔH. This can be seen clearly in Figure 6, which shows the temperature dependence of the average rmsd from the native structure and the average disulfide-bond content. For the

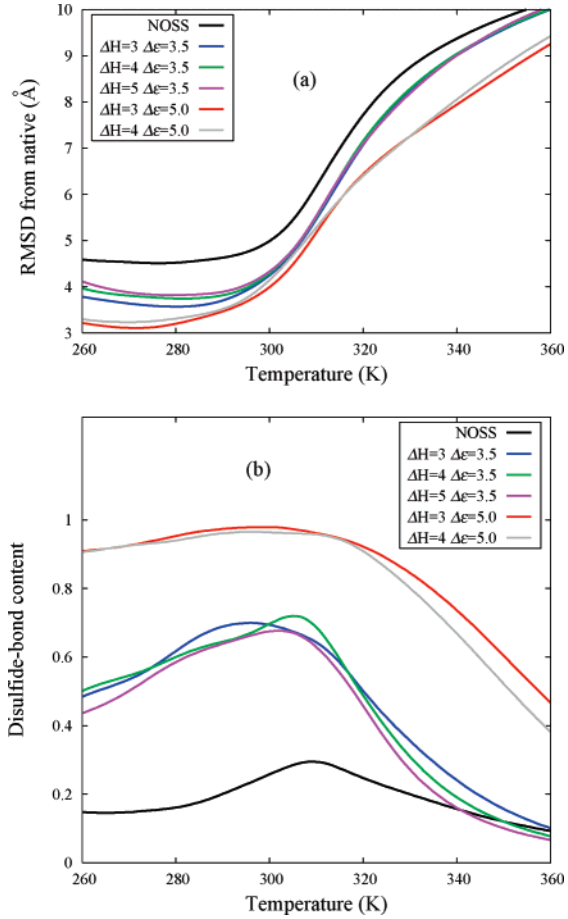

Figure 6. Temperature dependence of (a) the average rmsd from the native structure and (b) the average disulfide-bond content for protein 1ZDD, with data obtained from MREMD simulations. In the NOSS run, a disulfide bond is considered formed when the distance between Cys−Cys side-chain centroids is less than 8 Å.

purpose of calculating the disulfide-bond content in the NOSS run, a bond is considered formed when the distance between Cys−Cys side-chain centroids is less than 8 Å. The rmsd improves significantly as more disulfide bonds are present in the folded state (Figure 6), which is consistent with the data in Table 2. These results, particularly the low disulfide-bond content (less than 30%) for the NOSS run, imply that, without adding a real disulfide bond, the three-dimensional structure cannot accommodate even a half-cystine arrangement to readily form a disulfide bond, i.e., in order to form a disulfide bond, some structural modifications (with a possible introduction of strain) are necessary. For this reason, only with the higher value of $\Delta\epsilon$ is a folded state obtained, in which the bond is almost always formed.

The canonical MD simulations at 300 K show that this protein has no significant intermediates along the folding pathway. As seen in Figure 7(a), the time evolution of the fraction of folded structures (defined as those having an rmsd from the native structure below 6 Å) is almost unaffected by the introduction of the disulfide bond. In fact, the protein is already folded when the disulfide bond first appears in 70% of the cases for $\Delta H = 3$ and 80% for $\Delta H = 4$, regardless of $\Delta\epsilon$ (not shown here). The absence of an effect of ΔH is seen more clearly in Figure 7(b), which shows the

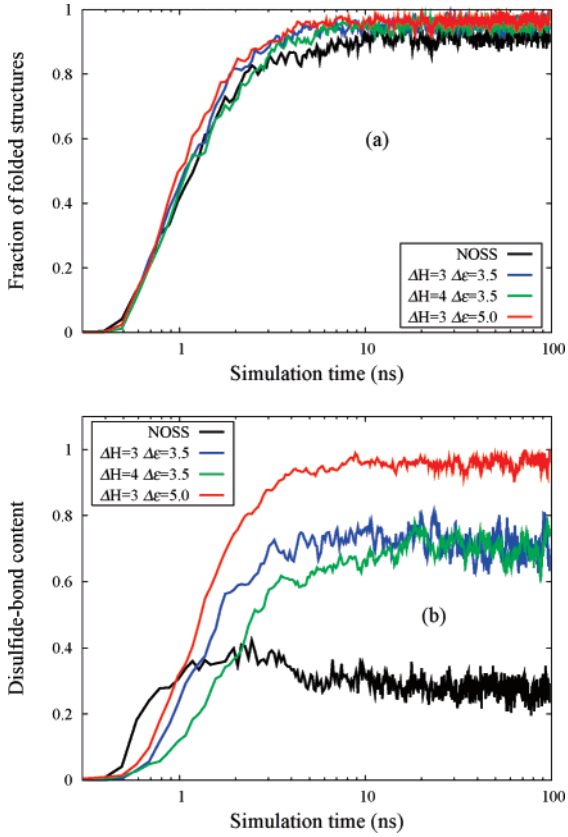

Figure 7. Time evolution of (a) the fraction of folded structures (rmsd from the native structure below 6 Å) and (b) the average disulfide-bond content in canonical MD simulations at 300 K for protein 1ZDD.

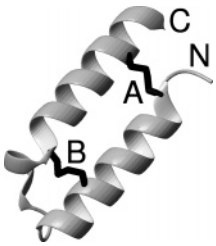

Figure 8. Native conformation[73] of 1EI0. The two disulfide bonds are highlighted in black and labeled A and B, and the N- and C-termini are labeled for clarity.

time evolution of the average disulfide-bond content. As expected, the two curves corresponding to $\Delta\epsilon = 3.5$ converge to the same value, but disulfide-bond formation is faster for lower ΔH (i.e., lower barrier), even though the final equilibrium value is independent of ΔH. It should be noted that these final values are also consistent with the results obtained from the MREMD simulations shown in Figure 6(b).

3.2. Protein 1EI0. 1EI0, shown in Figure 8, is a 38-residue protein with a fold very similar to that of 1ZDD. However, its structure[73] is stabilized by two disulfide bonds, between residues 3-34 (A) and 13-24 (B). The presence of four cysteine residues introduces the possibility of forming non-native disulfide bonds during folding. Experimental information[73,74] suggests that correct positioning of the helices precedes the formation of the disulfide bonds, which are known to form spontaneously. It was also shown that the

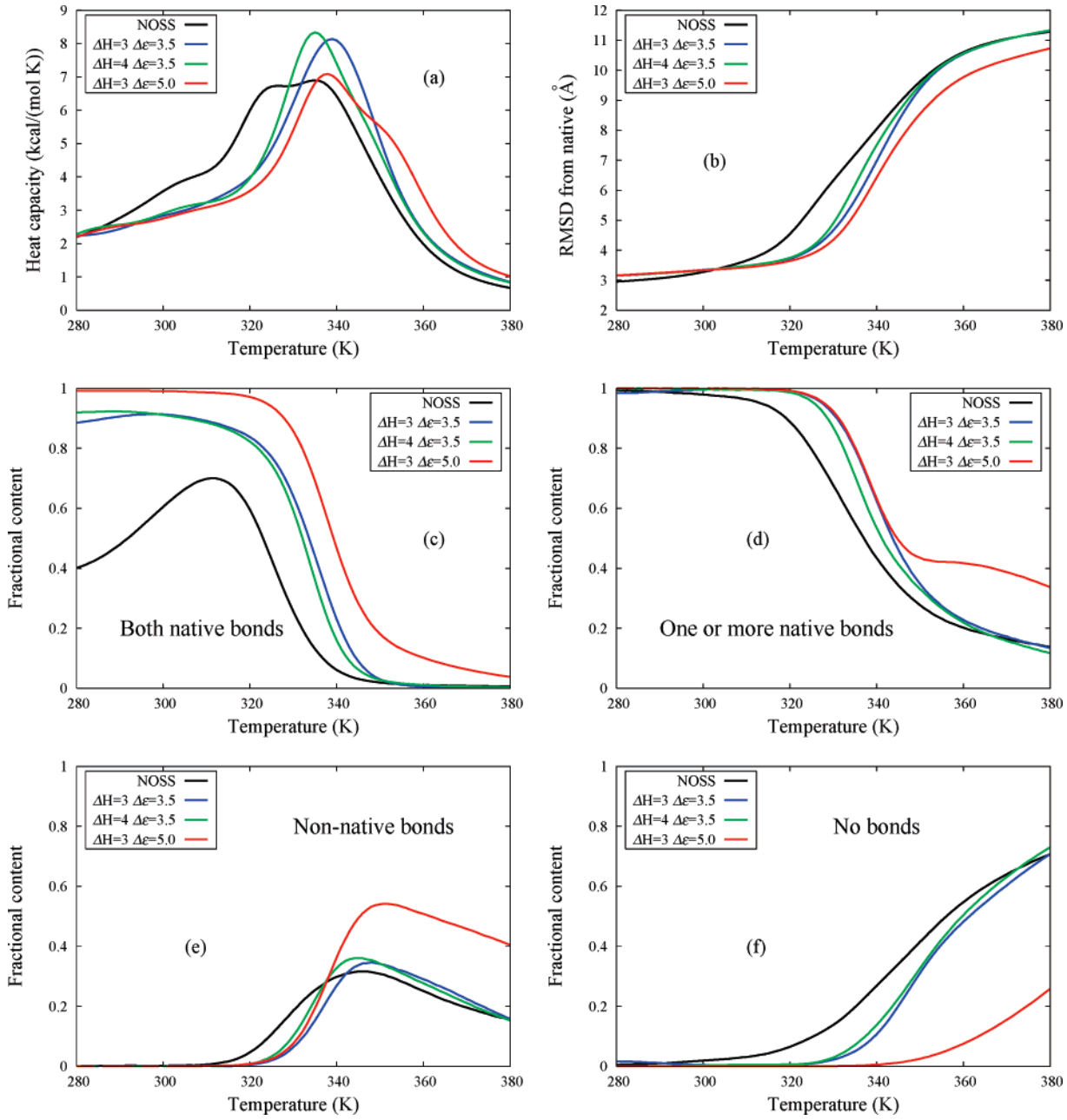

Figure 9. Temperature dependence of (a) the heat capacity and (b) the average rmsd from the native structure. Temperature dependence of the fraction of conformations with (c) both native disulfide bonds, (d) at least one native disulfide bond, (e) non-native disulfide bonds, and (f) no disulfide bonds. All plots refer to protein 1EI0, with data obtained from MREMD simulations.

corresponding peptide devoid of disulfide bonds is less stable but still has a high propensity to adopt a helical conformation.[73]

Figure 9 shows the results obtained with MREMD simulations. The folding transition occurs around 330−340 K, which is somewhat higher than for 1ZDD, and is sharper when disulfide-bond formation is allowed, as seen in Figure 9(a). Figure 9(b) shows that the results are in very good agreement with the experimental structure for all simulations (even NOSS), with average rmsd from the native structure just over 3 Å at 300 K. This protein has been studied previously with the UNRES force field, coupled with the CSA search method and dynamic disulfide-bond formation.[31] It was found in that study that the lowest-energy structure had only disulfide bond A of Figure 8 formed, even when using force field parameters designed especially for this protein. Figure 9(c),(d), on the other hand, shows that both native disulfide bonds are present in the majority of the conformations at 300 K, and virtually all conformations have at least one native bond [generally bond A (not shown here), as in the previous study[31]], even for the NOSS simulation. By contrast to 1ZDD, the introduction of disulfide bonds does not alter the conformations in the folded ensemble appreciably; the disulfide bonds can be formed with little strain and therefore lead to the formation of a high population of species with native disulfide bonds. While native disulfide bonds dominate the folded state (below the transition temperature), a significant number of structures containing non-native disulfide bonds (or no disulfide bonds at all) is found at higher temperatures [Figure 9(e),(f)]. It should be

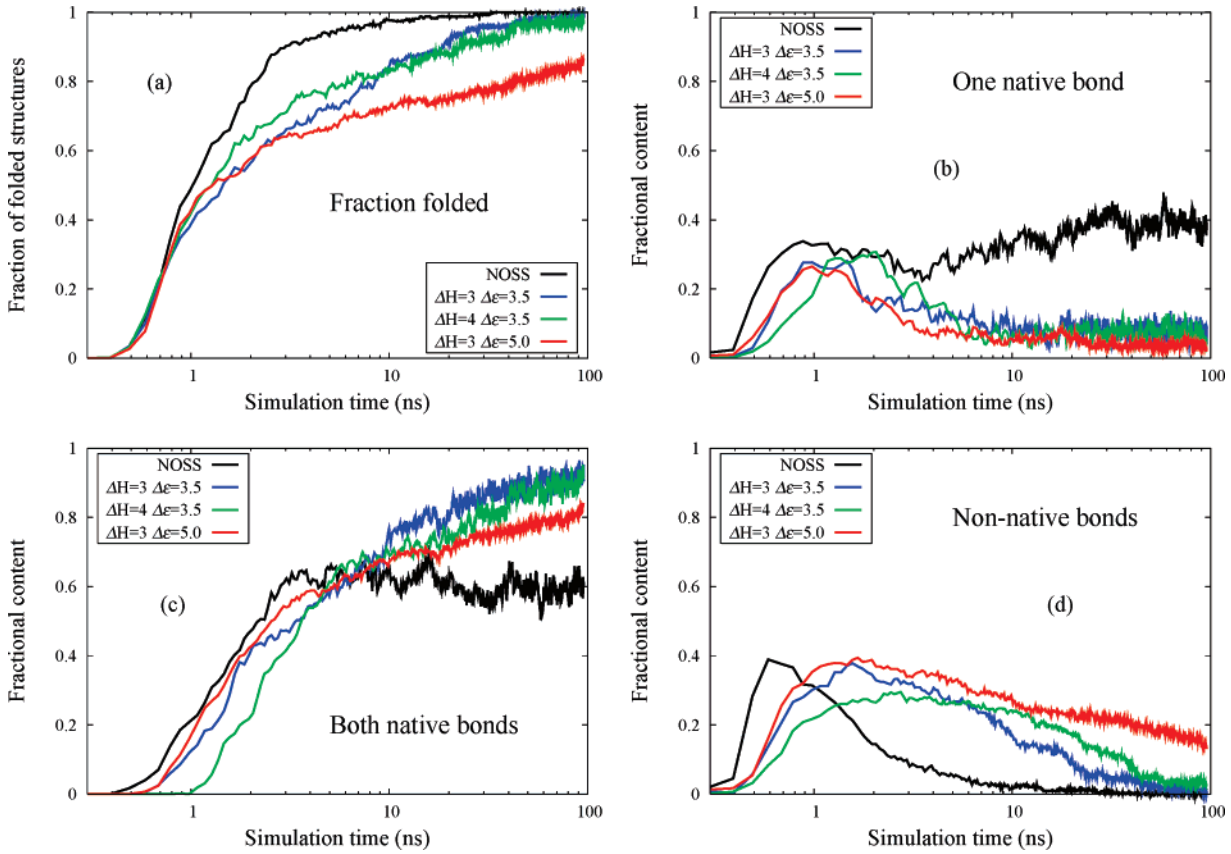

Figure 10. (a) Time evolution of the fraction of folded structures (rmsd from the native structure below 6 Å). Time evolution of the fraction of conformations with (b) one native disulfide bond, (c) both native disulfide bonds, and (d) non-native disulfide bonds. All plots refer to protein 1EI0, with data obtained from canonical MD simulations at 300 K.

noted that these conformations do not form any families with similar structure but rather a diverse collection of unfolded structures containing all possible disulfide-bond arrangements. As noted previously, the parameter ΔH has little effect on thermodynamic averages. On the other hand, the higher value of $\Delta\epsilon$ results in greater prevalence of disulfide bonds (native or otherwise) at higher temperatures.

Figure 10 shows the results obtained with canonical MD simulations at 300 K. By contrast to the results obtained for 1ZDD and shown in Figure 7(a), it is clear from Figure 10-(a) that folding is significantly slowed by the introduction of disulfide bonds. In fact, 100 ns were not sufficient to reach convergence for the run with $\Delta\epsilon = 5.0$. Disulfide bond B of Figure 8 generally forms before disulfide bond A (60−70% of the time, not shown here), but conformations with one native disulfide bond do not accumulate [Figure 10(b)] because the second disulfide bond forms readily [Figure 10-(c)]. Conformational folding in large part precedes the formation of fully-bonded structures, as seen from a comparison of Figure 10(a),(c) and from Table 3, which shows the percentage of structures that are already folded when various disulfide-bond arrangements first appear. The reason for the slower folding is the accumulation of structures containing non-native disulfide bonds [Figure 10(d)]. The NOSS run also visits these regions of conformational space, but only the presence of disulfide bonds turns them into kinetic traps, highlighting the downside of a large $\Delta\epsilon$. As mentioned in the previous paragraph, these structures do not represent a well-defined intermediate but rather a collection of misfolded conformations containing all possible single-bond arrangements (conformations with two non-native disulfide bonds never appeared).

Table 3. Percentage of Structures That Are Already Folded (rmsd from the Native Structure below 6 Å) When the Given Disulfide-Bond Arrangement (A, B, or A+B) First Forms in Canonical MD Runs at 300 K for Protein 1EI0

test system	A	B	A+B
NOSS	36	4	75
$\Delta\epsilon = 3.5$, $\Delta H = 3$	65	36	97
$\Delta\epsilon = 5.0$, $\Delta H = 3$	55	32	95
$\Delta\epsilon = 3.5$, $\Delta H = 4$	85	58	97

3.3. One-Disulfide Variant of 1GAB. The current force field can produce nativelike conformations for both 1ZDD and 1EI0, even without the introduction of disulfide bonds. While this is often the case for disulfide-stabilized proteins, there are examples of proteins for which the disulfide bonds are necessary for folding.[1] However, these proteins often contain β structure and are too complex to treat at present. Instead, two substitutions are made in the sequence of protein 1GAB, which was used to parametrize the force field used for this work.[50] 1GAB is a 47-residue, three-helix bundle.[75] In the new sequence, labeled 1GAB*, residues 9 and 26 (both Ala) were replaced by Cys. These sites were chosen because their side chains are not close in the native conformation, but they are close enough to form a disulfide bond in its mirror image; in our simulations here of 1GAB without

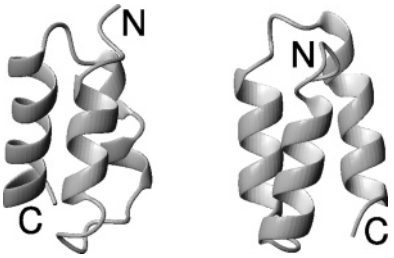

Figure 11. Native[75] (left) and mirror image (right) conformations of 1GAB. The N- and C-termini are labeled for clarity.

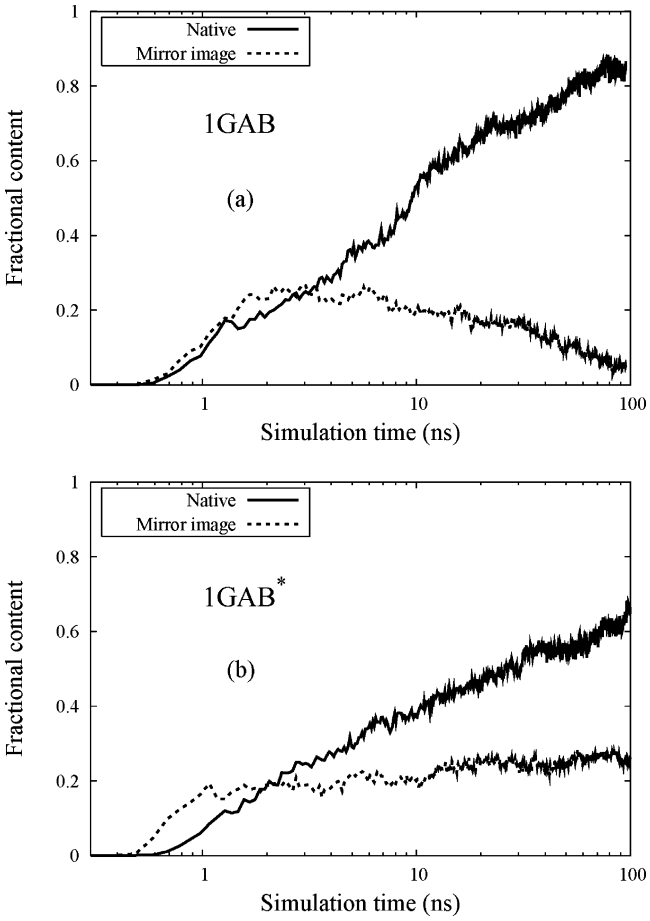

Figure 12. Time evolution of the native and mirror image populations in canonical MD simulations at 300 K for (a) 1GAB and (b) 1GAB* (with $\Delta H = 4$ and $\Delta\epsilon = 3.5$).

disulfide bonds, the mirror image appears as an intermediate during folding but is present only in less than 5% of the conformations in the equilibrium ensemble at 300 K, which is mainly nativelike. NOSS simulations of 1GAB* produce the same results, i.e., the effect of the introduction of Cys residues is negligible unless disulfide bonds are allowed to form. The native and mirror-image conformations are shown in Figure 11, while Figure 12(a),(b) shows the time evolution of the fraction of nativelike and mirror-image conformations (defined, as before, as those with rmsd within 6 Å from the corresponding native or mirror-image structure, respectively). In 1GAB, the mirror-image population grows as fast as the native population for the first 2 ns and subsequently decays slowly and almost disappears. In 1GAB*, on the other hand, the mirror-image population initially grows even faster than the nativelike population, but it never decays and in fact

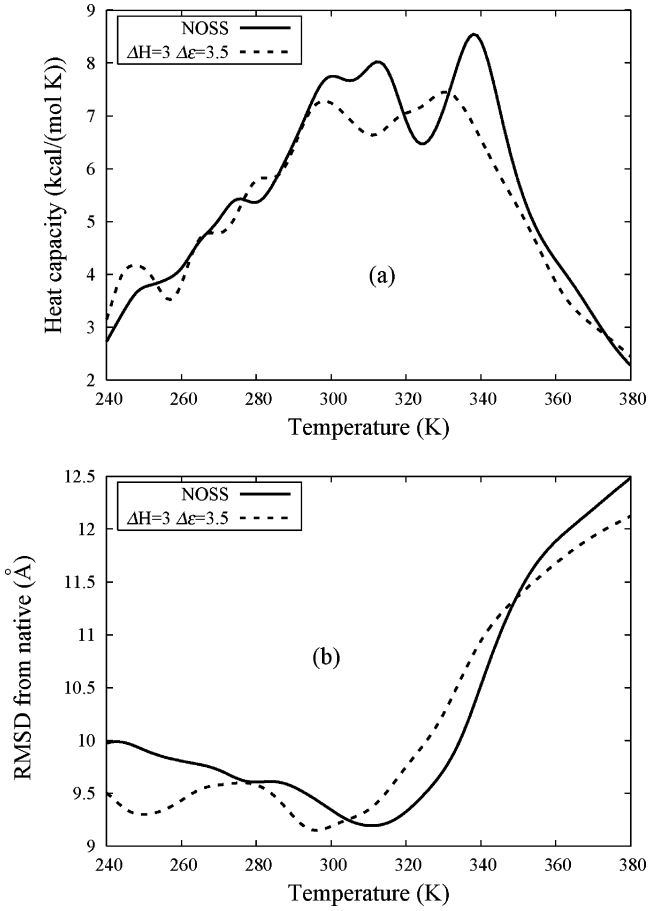

Figure 13. Temperature dependence of (a) the heat capacity and (b) the average rmsd from the native structure for protein 1NKL, with data obtained from MREMD simulations.

represents approximately 25% of the final state. Similar results for the final populations were obtained from MREMD simulations. Although we did not completely reverse the relative stabilities of the two main forms (native and mirror image), this result shows that it is possible to significantly alter the composition of the equilibrium state and not just increase its thermal stability as for 1ZDD and 1EI0.

3.4. Protein 1NKL. As a final test case, a larger and more complex protein was chosen, for which the current force field produces poorer predictions. 1NKL is a 78-residue four-helix bundle containing three disulfide bonds,[76] two of which staple the N-terminus to the C-terminus (4-76 and 7-70), while the third is in the middle of the sequence (35-45). The results of MREMD simulations are shown in Figure 13. In this case, the folding transition is not well defined, and the average rmsd of the predictions is very poor. The simulations were carried out for 200 ns (almost twice as long as for the other proteins) to ensure that lack of convergence was not causing the poor results. Although most predictions are far from native, the equilibrium state at 280 K (below the irregular transition region) does include a small fraction of nativelike conformations, as shown in Figure 13. These are present in both the NOSS run and in the run with dynamic disulfide-bond formation with $\Delta H = 3.0$ kcal/mol and $\Delta\epsilon = 3.5$ kcal/mol (which we label DYNSS) but are clearly more likely for the DYNSS run. In fact, the NOSS run does not produce any conformations with rmsd below 4 Å. We conclude that

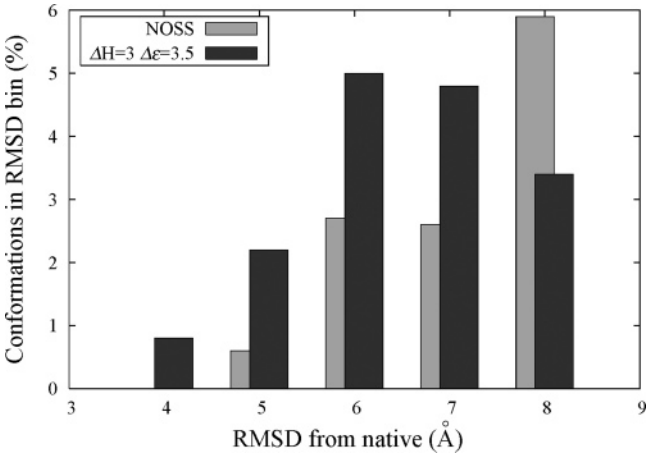

Figure 14. Percentage of conformations vs rmsd from the native structure for protein 1NKL, with data obtained from MREMD simulations.

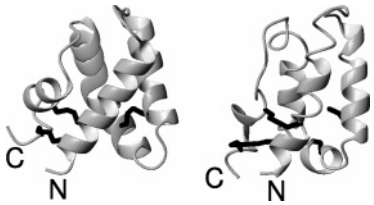

Figure 15. Native conformation[76] (left) and representative of the best family of predictions (right) for protein 1NKL. The disulfide bonds are highlighted in black, and the N- and C-termini are labeled for clarity.

the disulfide bonds again improve the quality of the predictions and the relative abundance of nativelike conformations but are not sufficient to turn them into the most probable family. When clustering is applied to the results of the MREMD simulations, a fragmented ensemble is revealed, reflecting the poorly defined transition. Approximately 20 families are found for each run, and, for both the NOSS and the DYNSS simulations, the third-ranking family is nativelike and represents 7−8% of the ensemble. However, in the NOSS run, this family has an average rmsd from the native structure of 6.2 Å, while disulfide bonds improve the rmsd of the corresponding DYNSS family to 5.0 Å. The native conformation and a representative from the nativelike DYNSS family (4.4 Å rmsd from the native structure) are shown in Figure 15. The predicted conformation contains the two native disulfide bonds near the termini, which appear to form much more easily than disulfide bond 35-45.

4. Conclusions

A new approach presented in this paper enables us to model dynamic formation and breaking of disulfide bonds in molecular dynamics simulations of protein folding. By using a reduced representation (UNRES), the time scale over which oxidative folding takes place becomes accessible. Disulfide-bond formation and breaking are simulated by introducing a transition barrier between the energy minima describing the disulfide bond on one side and the interaction between free half-cystine side chains on the other. To the best of our knowledge, this is the first algorithm based on physical principles to produce entire oxidative folding pathways for proteins, starting from the sequence alone and without using prior knowledge of the disulfide-bond arrangement.

The approach is tested on several helical proteins. Two of them (1ZDD and 1EI0) have a simple two-helix fold, with either one or two disulfide bonds. The force field produces nativelike conformations in both cases, even when dynamic formation of disulfides is not considered. However, when it is considered, the disulfide bonds are correctly predicted and the stability of the structure is improved, consistent with experimental findings. For protein 1ZDD, the quality of the prediction is also significantly improved. In both cases, many conformations are already nativelike by the time the native disulfide-bond arrangement is formed, a result which supports the folded-precursor mechanism of oxidative folding. However, this is not observed in all trajectories, suggesting that the quasi-stochastic mechanism also plays a role, albeit a smaller one for these proteins. Tests on a larger, more complex protein (1NKL) containing three disulfide bonds also support the conclusion that allowing the disulfide bonds to form results in improved predictions, with greater frequency and improved quality of nativelike conformations in the equilibrium ensemble. However, the nativelike family is not the dominant one in this case, and the results indicate that improvements are still necessary in the underlying UNRES force field (work currently under way). In a different kind of test on protein 1GAB, we show that mutation of the appropriate residues to cysteine can lead to overstabilized intermediate states, in which a disulfide bond is present that cannot form in the native conformation. Similar results have been observed experimentally.[8]

One possible problem with the potential function employed here is the formation of structures with more than two cysteines within bonding distance. Nothing in the model explicitely prevents this, but the results show that it is very rare, and therefore insignificant, because it is energetically unfavorable. However, an explicit term could be added to the potential to remove this possibility if it became a problem in future applications. Much more important for future work is the development of a more transferable force field that is able to treat more complex proteins, particularly those containing β structure. Several examples of such proteins exist for which oxidative pathways are known experimentally and would provide ideal test systems. The good results obtained here also suggest that it may be possible to use a similar approach to simulate other rare events in protein folding. One example is the cis−trans isomerization of proline residues.[77,78]

Acknowledgment. We thank Daniel Ripoll, Lovy Pradeep, and Robert Gahl for valuable suggestions and help with the selection of test proteins. This research was carried out by using the resources of our 820-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, of the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, of the National Center for Supercomputing Applications System at the University of Illinois at Urbana−Champaign, and of the Center for Computation and

Technology at Louisiana State University. All figures were prepared with the programs MOLMOL,[79] GNUPLOT,[80] and XFIG.[81]

References

(1) Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A. *Biochemistry* **2000**, *39*, 4207−4216.

(2) Shimotakahara, S.; Rios, C. B.; Laity, J. H.; Zimmerman, D. E.; Scheraga, H. A.; Montelione, G. T. *Biochemistry* **1997**, *36*, 6915−6929.

(3) Laity, J. H.; Lester, C. C.; Shimotakahara, S.; Zimmerman, D. E.; Montelione, G. T.; Scheraga, H. A. *Biochemistry* **1997**, *36*, 12683−12699.

(4) Starovasnik, M. A.; Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10081−10085.

(5) Creighton, T. E. *Science* **1992**, *256*, 111−114.

(6) Rothwarf, D. M.; Li, Y.-J.; Scheraga, H. A. *Biochemistry* **1998**, *37*, 3760−3766.

(7) Rothwarf, D. M.; Li, Y.-J.; Scheraga, H. A. *Biochemistry* **1998**, *37*, 3767−3776.

(8) Mason, J. M.; Cliff, M. J.; Sessions, R. B.; Clarke, A. R. *J. Biol. Chem.* **2005**, *280*, 40494−40499.

(9) Welker, E.; Wedemeyer, W. J.; Narayan, M.; Scheraga, H. A. *Biochemistry* **2001**, *40*, 9059−9064.

(10) Boudko, S. P.; Engel, J. *J. Mol. Biol.* **2004**, *335*, 1289−1297.

(11) Flory, P. J. *J. Am. Chem. Soc.* **1956**, *28*, 5222−5235.

(12) Poland, D. C.; Scheraga, H. A. *Biopolymers* **1965**, *3*, 379−399.

(13) Anfinsen, C. B.; Scheraga, H. A. *Adv. Prot. Chem.* **1975**, *29*, 205−300.

(14) Pace, C. N.; Grimsley, G. R.; Thomson, J. A.; Barnett, B. J. *J. Biol. Chem.* **1988**, *263*, 11820−11825.

(15) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Biochemistry* **1993**, *32*, 3178−3187.

(16) Betz, S. F. *Protein Sci.* **1993**, *2*, 1551−1558.

(17) Abkevich, V. I.; Shakhnovich, E. I. *J. Mol. Biol.* **2000**, *300*, 975−985.

(18) Zavodszky, M.; Chen, C.-W.; Huang, J.-K.; Zolkiewski, M.; Wen, L.; Krishnamoorthi, R. *Protein Sci.* **2001**, *10*, 149−160.

(19) Siadat, O. R.; Lougarre, A.; Lamouroux, L.; Ladurantie, C.; Fournier, D. *BMC Biochem.* **2006**, *7*, 12.

(20) Regan, L.; Rockwell, A.; Wasserman, Z.; DeGrado, W. *Protein Sci.* **1994**, *3*, 2419−2427.

(21) Rey, A.; Skolnick, J. *J. Chem. Phys.* **1994**, *100*, 2267−2276.

(22) Wang, Y.; Goh, S. Y.; Kuczera, K. *J. Pept. Res.* **1999**, *53*, 188−200.

(23) Qin, M.; Zhang, J.; Wang, W. *Biophys. J.* **2006**, *90*, 272−286.

(24) Camacho, C. J.; Thirumalai, D. *Proteins* **1995**, *22*, 27−40.

(25) Kobayashi, Y.; Sasabe, H.; Akutsu, T.; Saito, N. *Biophys. Chem.* **1992**, *44*, 113−127.

(26) Martelli, P. L.; Fariselli, P.; Malaguti, L.; Casadio, R. *Protein Eng.* **2002**, *15*, 951−953.

(27) O'Connor, B. D.; Yeates, T. O. *Nucleic Acids Res.* **2004**, *32*, W360−W364.

(28) Ferrè, F.; Clote, P. *Nucleic Acids Res.* **2005**, *33*, W230−W232.

(29) Ceroni, A.; Passerini, A.; Vullo, A.; Frasconi, P. *Nucleic Acids Res.* **2006**, *34*, W177−W181.

(30) Cheng, J.; Saigo, H.; Baldi, P. *Proteins: Struct., Funct., Bioinformatics* **2006**, *62*, 617−629.

(31) Czaplewski, C.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Protein Eng. Des. Sel.* **2004**, *17*, 29−36.

(32) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362−2367.

(33) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697−1714.

(34) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715−1731.

(35) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849−873.

(36) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874−887.

(37) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259−276.

(38) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(39) Lee, J.; Ripoll, D. R.; Czaplewski, C.; Pillardy, J.; Wedemeyer, W. J.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7291−7298.

(40) Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arłukowicz, P.; Ołdziej, S.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7299−7311.

(41) Liwo, A.; Arłukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1937−1942.

(42) Ołdziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035−8046.

(43) Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kozłowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421−9438.

(44) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16934−16949.

(45) Ołdziej, S.; Łagiewka, J.; Liwo, A.; Czaplewski, C.; Chinchio, M.; Nanias, M.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16950−16959.

(46) Ołdziej, S. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7547−7552.

(47) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13785−13797.

(48) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797 − 806.

(49) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100−1120.

(50) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260−285.

(51) Gay, J. G.; Berne, B. J. *J. Chem. Phys.* **1981**, *74*, 3316−3319.

(52) Kozlowska, U.; Liwo, A. To be submitted for publication.

(53) Lee, J.; Scheraga, H. A.; Rackovsky, *S. J. Comput. Chem*. **1997**, *18*, 1222−1232.

(54) Lee, J.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci*. U.S.A. **1999**, *96*, 2025−2030.

(55) Czaplewski, C.; Liwo, A.; Pillardy, J.; Ołdziej, S.; Scheraga, H. A. *Polymer* **2004**, *45*, 677−686.

(56) Doig, A. J.; Williams, D. H. *J. Mol. Biol.* **1991**, *217*, 389−398.

(57) Rothwarf, D. M.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 2671−2679.

(58) Rothwarf, D. M.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 2680−2689.

(59) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(60) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13798−13810.

(61) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637−649.

(62) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775−786.

(63) Czaplewski, C.; Kalinowski, S.; Scheraga, H. A. To be submitted for publication.

(64) Hansmann, U. H. E.; Okamoto, Y. *Physica A (Amsterdam)* **1994**, *212*, 415−437.

(65) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett*. **2000**, *329*, 261−270.

(66) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664−6675.

(67) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6676−6688.

(68) Nanias, M.; Czaplewski, C.; Scheraga, H. A. *J. Chem. Theory Comput*. **2006**, *2*, 513−528.

(69) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(70) Murtagh, F. *Multidimensional clustering algorithms*; Physika Verlag: Vienna, Austria, 1985.

(71) Murtagh, F.; Heck, A. *Multivariate data analysis*; Kluwer Academic: Dordrecht, Holland, 1987.

(72) Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5688−5692.

(73) Barthe, P.; Rochette, S.; Vita, C.; Roumestand, C. *Protein Sci.* **2000**, *9*, 942−955.

(74) Barthe, P.; Chiche, L.; Strub, M. P.; Roumestand, C. *J. Mol. Biol.* **1997**, *274*, 801−815.

(75) Johansson, M. U.; de Chateau, M.; Wikstrom, M.; Forsen, S.; Drakenberg, T.; Bjorck, L. *J. Mol. Biol.* **1997**, *266*, 859−865.

(76) Liepinsh, E.; Andersson, M.; Ruysschaert, J. M.; Otting, G. *Nat. Struct. Biol.* **1997**, *4*, 793−795.

(77) Boulegue, C.; Milbradt, A. G.; Renner, C.; Moroder, L. *J. Mol. Biol.* **2006**, *358*, 846−856.

(78) Pradeep, L.; Shin, H. C.; Scheraga, H. A. *FEBS Lett.* **2006**, *580*, 5029−5032.

(79) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.

(80) http://www.gnuplot.info.

(81) http://www.xfig.org.

CT7000842