An Efficient, Path-Independent Method for Free-Energy Calculations

Michael D. Tyka,* Anthony R. Clarke, and Richard B. Sessions

Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, United Kingdom

Received: February 3, 2006; In Final Form: June 21, 2006

Classical free-energy methods depend on the definition of physical or nonphysical integration paths to calculate free-energy differences between states. This procedure can be problematic and computationally expensive when the states of interest do not overlap and are far apart in phase space. Here we introduce a novel method to calculate free-energy differences that is path-independent by transforming each end state into a reference state in which the vibrational entropy is the sole component of the total entropy, thus allowing direct computation of the relative free energy. We apply the method to calculate side-chain entropies of a β-hairpin-forming peptide in a variety of backbone conformations, demonstrating its importance in determining structural propensities. We find that low-free-energy conformations achieve their stability through optimal trade off between enthalpic gains due to favorable interatomic interactions and entropic losses incurred by the same.

1. Introduction

Free-energy calculations of biomolecular systems have been the subject of intense research over the past decades.[1−4] In some cases this problem can be stated as a transformation of the system of interest from one Hamiltonian to another where the two Hamiltonians may, for example, represent different ligands in the same binding site. For these cases methods such as thermodynamic integration and free-energy perturbation have been increasingly successful in the calculation of free-energy differences. Classic applications of these methods include the calculation of solvation free energies,[5−8] relative and absolute ligand-binding affinities,[9−12] and free-energy differences associated with amino acid mutations[11] in which the states of interest are alchemically transformed in different environments. The states of interest rarely overlap sufficiently in phase space to calculate free-energy differences directly, and thus the transformation is usually subdivided into steps along a pathway of intermediates. Since free energy is a state function, the free-energy difference between two states is independent of the path, and thus the intermediates do not have to be physically realistic. Much of the power of free-energy calculations derives from this fact. However, the choice of pathway is critical in practice since some pathways allow the calculations to converge considerably faster than others. When the states of interest are far apart in phase space or overlap poorly, finding a suitable integration path can be difficult, and as more intermediate states are required, more simulation time is needed to collect the necessary data. In some situations the usage of soft-core potentials or the definition of a judiciously chosen reference state can overcome some of these problems;[12] however, the latter has to be designed specifically for any given problem. Other techniques calculate free-energy changes along some reaction coordinate, and various schemes have been proposed to enhance sampling along the coordinate path.[13−15] Such methods have been successfully applied to conformational studies[16,17] and ligand binding,[18−20] for example. However, if the end states are separated by a long path and the phase space perpendicular to the reaction coordinate between the end states of interest is too extensive, then the system can become trapped despite the presence of biasing potentials, since these only flatten the energy landscape in the direction of the reaction coordinate.[17]

In this paper, we present a new method for calculation of free-energy differences of states that does not rely on construction of a pathway between them. Instead, each state of interest is transformed into a harmonically confined reference state such that the free-energy differences between the original states and the reference states as well as between the two reference states can be calculated directly. Harmonic restraints have been widely used in a variety of different ways to overcome sampling problems in free-energy calculations. Examples include enhancement of sampling in regions of high free-energy (umbrella sampling[13,14]) and, during the calculation of association constants, providing a restrained intermediate state to prevent sampling problems when the host system and the ligand become fully decoupled.[21−25] Here we use harmonic restraints to freeze the system in the states of interest in a manner similar to Štrajbl et al.[26] and, in combination with a normal-mode calculation, show how free-energy differences can be calculated without the need to transform the states of interest into each other.

First we test this new method on a model system small enough to allow direct comparison with a free-energy perturbation approach and demonstrate that the two methods yield the same result, although even for such a small test system the confinement method requires much less computational effort. Second we apply the method to a larger system for which a direct path method would not be computationally feasible: The example application that we use is the calculation of the relative free energy and the side-chain entropy of a peptide in several, highly diverse conformations. Residue-dependent entropy and enthalpy changes have been proposed to play a major role in determining the secondary structure propensity of different amino acid sequences,[27,28] and thus it is of interest to study the free-energy and entropy differences of peptides as a function

* Author to whom correspondence should be addressed. E-mail: m.tyka@bris.ac.uk.

10.1021/jp060734j CCC: $33.50
Published on Web 08/08/2006

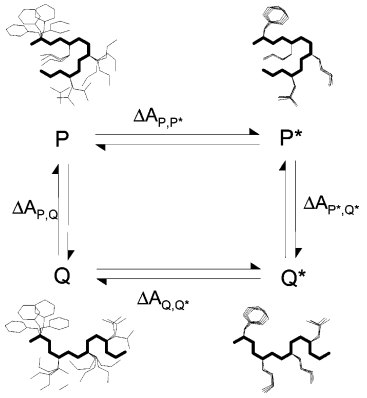

Figure 1. Thermodynamic cycle illustrating the confinement method. The free and restrained states (the latter is marked with a asterisk) overlap sufficiently in phase space to allow the free-energy differences $\Delta A_{P,P^*}$ and $\Delta A_{Q,Q^*}$ to be calculated using thermodynamic integration. Although the two restrained states P* and Q* have no overlap in phase space, $\Delta A_{P^*,Q^*}$ can be obtained accurately from two normal-mode calculations. Thus this approach allows the calculation of $\Delta A_{P,Q}$ in a path-independent manner since it eliminates the need to transform P into Q at any stage of the algorithm.

of backbone conformation in detail to better understand the sequence−structure relationships of proteins.

Free energies of peptides are usually calculated using direct simulation and projection of the free energy onto one or more order parameters, structure-dependent quantities such as the radius of gyration, the root-mean-square deviation (RMSD) to the native structure, or the number of native hydrogen bonds.[29−32] However, projection of the multidimensional configuration space of a peptide onto a small number of order parameters oversimplifies the multidimensional free-energy surface and gives a false impression of a relatively smooth folding landscape.[33] Further, because a large number of different peptide conformations can share the same set of order parameters it does not allow dissection of the landscape into free energies of individual conformations nor does it reveal much information about the structural origins of free-energy or entropy differences between conformations.

Entropic differences between systems can also be estimated directly using normal-mode analysis[34] or quasi-harmonic (QH) analysis.[35,36] Calculation of normal modes of a peptide, which approximate the energy surface around a minimum to be harmonic, will account for differences in vibrational entropy but cannot account for differences in configurational entropy. In an approach proposed recently[37,38] this limitation is overcome by including *every* minimum in a given state, calculating its potential energy and vibrational entropy, and combining the results of all the relevant minima into an estimate of the total free energy. However, this becomes increasingly difficult with larger peptides as the number of individual minima becomes astronomical, and thus application to larger peptides has not been feasible. Quasi-harmonic analysis estimates entropy differences by approximating the atomic fluctuations around the average particle positions by a Gaussian probability distribution.[35,39] Although early results suggested that this appears to be a good approximation in condensed systems,[40] recent work has shown that its validity depends strongly on the type of entropy to be calculated.[41,42]

2. Method

2.1. Approach. The approach to the calculation of the free-energy difference of two states without phase space overlap is illustrated in Figure 1. We seek to calculate $\Delta A_{P,Q}$, the free-energy difference between two states P and Q. Instead of transforming P into Q, as would be necessary in classical approaches, we transform each into a frozen state, P* and Q*, respectively, by applying a harmonic restraint potential to all atoms. The free energy of this process, $\Delta A_{P,P^*}$ and $\Delta A_{Q,Q^*}$, can be calculated using a classical approach, since the free and the frozen states overlap well in phase space. Here we chose to use thermodynamic integration[2] for this purpose. Finally, to close the thermodynamic cycle shown in Figure 1, we must calculate $\Delta A_{P^*,Q^*}$. Because the atoms in the frozen states are harmonically restrained the only remaining contributor to the entropy of these states is the vibrational entropy. All other contributions such as rotational entropies (e.g., alternative rotameric states of side chains) will have been frozen out in the restraint process. This allows us to calculate the free-energy difference $\Delta A_{P^*,Q^*}$ analytically using normal-mode analysis on P* and Q*. Adding the free-energy differences around the thermodynamic cycle thus allows us to calculate $\Delta A_{P,Q}$ without the need to transform P into Q at any stage of the algorithm. In essence the normal-mode calculation is used as a shuttle to jump from one part of conformational space to another, independent of the distance or intervening barriers. In fact, the two end states P and Q could equally belong to an entirely different Hamiltonian or system.

A related approach was used by Štrajbl and Warshel[26,43] to calculate the entropic contribution to the activation energies of chemical reactions in which motion perpendicular to the reaction path was harmonically restrained and the associated free-energy cost evaluated by free-energy perturbation. The method presented here generalizes this approach by freezing out *all* the motions, which, combined with the normal-mode calculation, allows evaluation of free-energy differences between distant states.

We apply the new method first to a small pentapeptide, Met-enkephalin, showing that the results are in agreement with a free-energy perturbation approach, and then apply it to a larger, 17 amino acid fragment of ubiquitin.

2.2. Theory. Thermodynamic integration[2,44] expresses the free-energy difference ΔA between two states as

$$\Delta A = \int_0^1 \left\langle \frac{\partial E(\mathbf{X},\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \qquad (1)$$

where λ is a dimensionless parameter defining the integration path running from 0 to 1, $\partial E(\mathbf{X},\lambda)/\partial\lambda$ is the derivative of the potential energy with respect to λ, and $\langle\cdots\rangle_\lambda$ signifies an ensemble average at a particular λ. This integration is usually performed numerically by recording individual values of $\partial E(\mathbf{X},\lambda)/\partial\lambda$ and using the trapezoidal or Simpson's rule.

To calculate the free-energy change of the freezing process as outlined above we need to perturb the free system P at $\lambda = 0$ to the restrained system P* at $\lambda = 1$. The energy function $E(\mathbf{X},\lambda)$ is thus

$$E(\mathbf{X},\lambda) = E_{\mathrm{ff}}(\mathbf{X}) + (^1/_2)\lambda k_{\mathrm{f}}|\mathbf{X} - \mathbf{X}_0|^2 \qquad (2)$$

where E_{ff} is the force field energy, $\mathbf{X}$ is the vector of atomic positions, $\mathbf{X}_0$ is their original position, and k_{f} is the final restraint force constant at $\lambda = 1$. The second term is the restraint used to freeze the system. Here we chose to use an absolute Cartesian restraint, although in principle any type of restraint (e.g., an internal restraint) is suitable provided it is sufficiently strong at $\lambda = 1$ such that the normal-mode approximation applies well.

If we write the particular k at which a given simulation takes place as $k = \lambda k_{\mathrm{f}}$, then we can restate the above integral using k as the variable of integration instead of λ

$$\Delta A = \int_0^1 \left\langle \frac{\partial E(\mathbf{X},\lambda)}{\partial \lambda} \right\rangle d\lambda = (^1/_2)\int_0^{k_f} \langle |\mathbf{X} - \mathbf{X}_0|^2 \rangle_k \, dk = (^1/_2)\int_0^{k_f} \chi_k \, dk \quad (3)$$

since

$$\frac{\partial E(\mathbf{X},\lambda)}{\partial \lambda} = (^1/_2)k_f|\mathbf{X} - \mathbf{X}_0|^2 \quad \text{and} \quad d\lambda = (1/k_f)\, dk$$

This enables one to simply record the ensemble averages χ_k of the atomic deviations at different values of k and then perform a numerical integration to obtain a free-energy difference.

2.3. Numerical Integration. Figure 2 shows a typical data set to be integrated where the area underneath the curve will be equal to the free-energy difference. The simple trapezoidal rule for numerical integration interpolates linearly between successive points, clearly a bad approximation since there is a strong curvature. Replotting the data in a double log plot shows that a significant proportion of the graph fits well to a function of the form $\chi_k = ak^b$ (where χ_k designates the ensemble average $\chi_k = \langle |\mathbf{X} - \mathbf{X}_0|^2 \rangle_k$, see eq 3). Fitting such a function between successive points greatly improves the accuracy of the numerical integration and allows fewer integration points to be used, thus increasing the computational efficiency. Given two successive points of data $P_i = (\chi_{ki}, k_i)$ and $P_j = (\chi_{kj}, k_j)$ with $j = i + 1$, the area L_i underneath the fitted curve is

$$L_i = \frac{a}{1+b}(k_j^{b+1} - k_i^{b+1}) \quad (4)$$

where

$$a = \exp[\ln(\chi_{ki}) - b \ln(k_i)] \quad (5)$$

and

$$b = \frac{\ln(k_j) - \ln(k_i)}{\ln(\chi_{kj}) - \ln(\chi_{ki})} \quad (6)$$

Thus the total free-energy change between states P and P*, $\Delta A_{P,P^*}$, is obtained from

$$\Delta A_{P,P^*} \cong (^1/_2)(\chi_0 + \chi_{k1})k_1 + \sum_{i=1} L_i \quad (7)$$

where the contribution to the free-energy change between $k = 0$ and $k = k_1$ is calculated using the trapezoidal rule and the remainder is obtained using the improved fitting procedure (eqs 4−6).

2.4. Normal-Mode Analysis. The free energy of the restrained conformation is calculated using the classical formula for the canonical partition function of a system of N particles with κ vibrational degrees of freedom[45]

$$Z = \exp(-E_0/k_B T)\prod_i^{\kappa} \frac{k_B T}{h\upsilon_i} \quad (8)$$

where υ_i is the ith normal-mode frequency, E_0 is the potential energy at the minimum, T is the temperature, h is Plank's constant, and k_B is the Boltzmann constant. Note that we use the classical expression for the partition function rather than the quantum mechanical one because only then will the overall result for the free-energy change converge with increasing k_f.

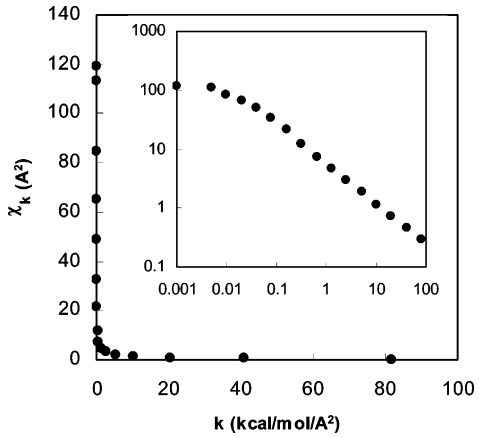

Figure 2. Typical dataset obtained from the thermodynamic integration. The inset illustrates the excellent fit to a function of the form $\chi_k = ak^b$.

Knowledge of the partition function enables one to calculate the free energy of the restrained state (A_{P^*}) directly

$$A_{P^*} = -k_B T \ln(Z) \quad (9)$$

Thus the remaining leg of the thermodynamic cycle shown in Figure 1 can be computed directly without the need to construct a path to transform P into Q

$$\Delta A_{P^*,Q^*} = A_{Q^*} - A_{P^*}$$

The $3N$ normal-mode frequencies are obtained in the usual manner, by finding the eigenvalues of the mass-weighted matrix of second derivatives (Hessian) of the potential energy at the minimum using standard procedures. Here the Hessian is calculated through numerical differentiation of the forces since it was not a time-critical part of the calculation. Note that since the restraints are absolute with respect to the coordinate system the system has $3N$ instead of $3N - 6$ degrees of freedom.

2.5. Summary of the Method. The calculation of the free energy of a peptide is thus performed as follows. First, the backbone of the peptide is restrained by a harmonic potential to define the configurational state. Then, to calculate the first leg of the thermodynamic cycle a number of simulations are set up, each with an additional harmonic restraint applied to all atoms and with an increasing force constant. At each simulation the average total squared atomic deviation from the original restraint position is calculated. The data are then numerically integrated using the modified fitting technique described above. At the final restraint strength a normal-mode analysis is performed, and the classical entropy is calculated. If this process is repeated for each state of interest, then the free-energy difference ΔA_{PQ} (Figure 1) can be obtained

$$\Delta A_{PQ} = \Delta A_{P,P^*} - \Delta A_{Q,Q^*} + \Delta A_{P^*,Q^*} = (A_{Q^*} - \Delta A_{Q,Q^*}) - (A_{P^*} - \Delta A_{P,P^*}) \quad (10)$$

Note that the result is independent of the nature and strength of the confining restraint potential provided it is sufficiently strong to obtain $\Delta A_{P^*,Q^*}$ accurately; i.e., with increasing final restraint strength, k_f, the values obtained for $\Delta A_{P,Q}$ converge as we will show in the Results section (see subsection 3.1 and Figure 4).

2.6. Application to the Calculation of Side-Chain Entropies. We show how the above method can be used to rigorously calculate side-chain entropies of small peptides in various conformations. To remove entropy contributions of the backbone we apply a moderate Cartesian restraint to all backbone atoms

during all of the simulations. The backbone can be fixed to different conformations in turn allowing changes in side-chain entropies to be calculated. Note that this restraint is completely unrelated to the confinement restraint used to calculate the free-energy differences between states as detailed in subsections 2.1−2.3. The backbone constraint is simply added to the force field energy E_{ff}

$$E_{ff} = E_{MM} + (^1/_2)k_{bb}|\mathbf{B} - \mathbf{B}_0|^2$$

where E_{MM} is the molecular mechanics potential energy calculated using a standard force field (here we use AMBER ff03[46]), k_{bb} is the force constant of the backbone restraint applied to the vector $\mathbf{B}$ of backbone atoms, and $\mathbf{B}_0$ is their restraint position defining the conformation. Here we use a backbone constraint of 1.0 kcal/mol/Å^2. Because each respective conformation is fixed in a different position in phase space there is no phase space overlap between the different states, and thus the calculation of free-energy differences between these states using classical approaches such as free-energy perturbation would be computationally inefficient since the integration path necessary would be very long. One would have to adiabatically morph one set of backbone constraints into the other along an unfolding and refolding path, calculating free-energy differences for each small step. Although we will use such an approach on a small model system to demonstrate its equivalence to the newly developed method shown here, the free-energy perturbation method is computationally very expensive and not applicable to anything larger than a very small model system.

2.7. Validation on a Small Test System. To validate the proposed method for free-energy calculation we compare the results obtained using the confinement approach with an established method; here we use free-energy perturbation (FEP). The model system must be sufficiently small such that a path-dependent method such as FEP can be used accurately. Here we chose Met-enkephalin, a small pentapeptide. For the purpose of validation we keep the simulation complexity to a minimum by neglecting proper treatment of solvent. Met-enkephalin (NH_3-Tyr-Gly-Gly-Phe-Met-COO) was simulated using the AMBER ff03 force field[46] and a distance-dependent dielectric constant[47] $\epsilon(r) = 1 + r$. Seven structurally diverse, low-energy conformations were chosen from a preliminary conformational space annealing (CSA)[48] simulation and will be referred to as Met_0−Met_6. All simulations were run using Langevin dynamics[49] ($\gamma = 1.0$ ps^{-1}) implemented into our in-house simulation software Boa.

Free energies were calculated using the confinement method using the protocol of simulations shown in Table 1. The integration time step was reduced at higher confinement force constants, since a higher accuracy was required when the confinement was strong. Equally, the tighter the confinement, the faster the simulations converge, and thus much shorter simulations can be used. Simulations at force constants 0.001−0.640 kcal/mol/Å^2 were simulated using replica exchange molecular dynamics (REMD)[50] to ensure thorough sampling. Ten replicas were used at temperatures exponentially arranged from 290−540 K, each replica was simulated for $^1/_{10}$th of the number of steps shown in Table 1. The remainder of the simulations were run using a single temperature run at 300 K because at higher restraint strengths the system essentially remains in one conformational state and REMD becomes disadvantageous since sampling time is wasted on high-temperature simulations that are of no interest here. The first 40% of each simulation was allowed for equilibration, and thermodynamic averages were calculated only over the remainder of the simulation. Preliminary tests showed that this arrangement gave the best convergent properties. Errors were estimated by dividing the last 60% of each trajectory into three blocks and calculating the free energies separately. The final k, k_f, has to be sufficiently high such that the harmonic approximation applies well; here we simulate up to $k_f = 81.92$ kcal/mol/Å^2. Note that that we approximated χ_{k0} with χ_{k1} instead of running another simulation at $k = 0$.

TABLE 1: Thermodynamic Integration Protocol

k (kcal/(mol/Å^2))	steps ($\times 10^3$)	type	time step (fs)
0.001	2000	REMD	1.0
0.005	2000	REMD	1.0
0.01	2000	REMD	1.0
0.02	2800	REMD	1.0
0.04	2800	REMD	1.0
0.08	2800	REMD	1.0
0.16	2800	REMD	1.0
0.32	2400	REMD	1.0
0.64	2000	REMD	1.0
1.28	800	MD	1.0
2.56	600	MD	1.0
5.12	400	MD	1.0
10.24	400	MD	0.5
20.48	400	MD	0.5
40.96	400	MD	0.5
81.92	400	MD	0.5

For validation, free energies were also calculated using FEP. The free-energy difference between two conformations, say m and n, was calculated by slowly perturbing one set of back backbone restraint positions B_0^m into another, B_0^n, along a path that slowly morphs peptide conformation m into peptide conformation n. The morph path was created by linearly interpolating the backbone angles of the two conformations in 134 steps (including the endpoints) to avoid excessively large free-energy differences at any one step. The perturbation was performed in both directions with the hysteresis between the two results indicating the amount of uncertainty in the result. The seven different conformations were "connected" by six morph paths, allowing the calculation of all relative free-energy differences. Some paths were found to cause severe steric clashes, and thus a suitable set of paths was chosen. These paths were from Met_0 to Met_1, Met_0 to Met_2, Met_0 to Met_3, Met_1 to Met_4, Met_2 to Met_5, and Met_2 to Met_6. To ensure convergence, 24 ns of Langevin dynamics was performed at each step for a total of ∼3.2 μs of simulation.

2.8. Calculation of Side-Chain Entropies of a 17-mer. U(1−17)T9D is a mutant 17-residue peptide fragment (MQIFVKTLDGKTITLEV), derived from ubiquitin and shown to fold independently into a hairpin structure.[51] This peptide was simulated using the AMBER03 force field[46] and an implicit solvation potential, GB/SA.[52,53] The nonpolar solvation energy coefficient[54] was set to 0.005 kcal/mol/A^2. Electrostatic interactions were cut off at 15 Å (using a switching function between 13 and 15 Å) while van der Waals interactions were tapered off to 0 from 4.5 to 5.5 Å. The latter cutoffs were introduced to reduce the overestimation of intrasolute van der Waals interactions relative to solute−solvent interactions, which are not dealt with appropriately using surface-area-based models.[55]

Twelve different, low-potential-energy conformations of U(1−17)T9D were chosen from a larger set generated using CSA[48] and a novel β-sheet-generating Monte Carlo move developed in our laboratory (M.D.T., A.R.C., unpublished results). The conformations were chosen by inspection to include a large structural variety: a near native structure with a distorted hairpin, two hairpins with incorrect strand register, several

TABLE 2: Results of Free-Energy Calculations on Conformations of U(1−17)T9D[a]

structure	type	RMSD	ΔA	ΔU	$-T\Delta S$
Ubi_01	β-hairpin, native	1.2	-18.6 ± 1.1	-39.1	20.5
Ubi_02	β-hairpin, native	1.5	-21.7 ± 1.2	-45.1	23.4
Ubi_03	β-hairpin, native-like, distorted turn	2.2	-9.3 ± 0.3	-29.7	20.4
Ubi_04	β-hairpin, non-native strand register	2.7	-14.1 ± 0.5	-29.8	15.8
Ubi_05	β-hairpin, non-native strand register	5.1	-13.6 ± 1.2	-33.7	20.1
Ubi_06	partially α-helical	10.8	-4.7 ± 1.5	-24.8	20.1
Ubi_07	α-helical, compact	5.8	-16.9 ± 1.2	-42.4	25.5
Ubi_08	α-helical, noncompact	7.4	-19.8 ± 0.8	-42.7	22.8
Ubi_09	partially α-helical	8.8	-15.6 ± 0.8	-38.7	23.1
Ubi_10	α-helical, compact	6.3	-14.3 ± 1.6	-41.0	26.6
Ubi_11	α-helical, compact	5.8	-7.8 ± 1.4	-42.1	34.3
Ubi_12	α-helical, compact	6.0	-12.1 ± 1.4	-39.2	27.1
Ubi_13	full α-helix	10.1	-14.1 ± 0.6	-34.2	20.1
Ubi_14	fully extended	16.6	0.0 ± 1.7	0.0	0.0

[a] Calculated free energies (ΔA), internal energies (ΔU), and entropies (given as $-T\Delta S$, with $T = 300$ K) are shown in kcal/mol relative to Ubi_14. Column 3 shows the root-mean-square distance deviation of the C_α to the published structure of U(1−17)T9D.

partially helical structures, one full helix, and a fully extended structure. Additionally, two further structures were generated by optimizing the native structure (PDB code 1E0Q, model 1), resulting in two structures at 1.2 and 1.5 Å C_α RMSDs, respectively. The 14 structures (referred to as Ubi_01 to Ubi_14), their approximate conformations, and RMSDs from the published native state are summarized in Table 2. Internal energies, U, were calculated as average potential energies from a single 10 ns simulation without further constraints, except the backbone constraint. Entropies were calculated by using the relationship $\Delta S = (\Delta U - \Delta A)/T$, with $T = 300$ K.

2.9. Scope and Limitations. The method presented is universally applicable in principle, as long as one can confine the system to a state in which the vibrational entropy is the only significant contributor to the total entropy. Note that the computational effort is only dependent on the extent of conformational freedom (the volume of phase space) in the unconfined end states but is independent of the length of any intervening path. Thus the method should prove superior to path-dependent methods in situations where the end states are radically different from each other, for example, ligand-binding reactions accompanied by large conformational changes. The approach is also computationally well suited to parallel computer architectures because each simulation can be run separately and REMD simulations are easily parallelizable themselves, requiring very little inter-processor communication.

When used in the context of explicit solvent, any method for free-energy calculation must take into account any changes in the solvent enthalpy and entropy. In the present case, this could be achieved by applying the confinement to all atoms of the system including the solvent molecules. In this way the solvation energy including the entropy would be taken into account in principle. However, accurately calculating the free energy of confining the solvent molecules to one position in the solvent box would require their average distance from their restraint position to converge. At low restraint strengths this cannot be expected in a computationally feasible simulation. A similar problem arises with ligand-binding free-energy calculations when the ligand is decoupled from the remainder of the system and thus free to diffuse away from the binding site.[10,25] Consequently, the algorithm presented in its present form is limited to simulations performed in implicit solvent models, although current work in our laboratory is addressing this limitation.

3. Results and Discussion

3.1. Results for Met-Enkephalin. Figure 3 shows that the free-energy differences calculated by the confinement method

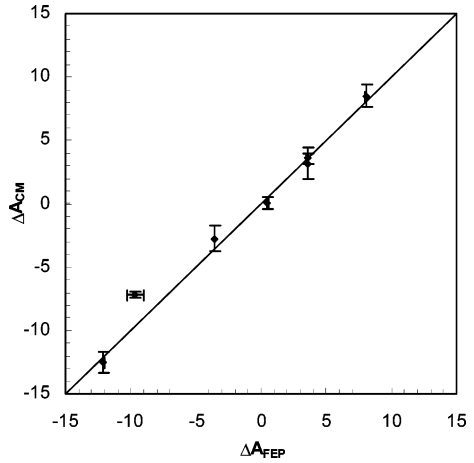

Figure 3. Free energies of different conformations of Met-enkephalin calculated using the confinement method (ΔA_{CM}) and using free-energy perturbation (ΔA_{FEP}) in kcal/mol. Vertical error bars indicate estimated errors for the confinement method. Horizontal error bars show discrepancy between forward and backward FEP calculations, which are only appreciable in the second point from the left due to lack of overlap in the FEP integration path used for that conformation (see text).

are in excellent agreement with those calculated by a direct path perturbation method. Note that the forward and backward perturbation calculations are in excellent agreement with each other (discrepancies <0.2 kcal/mol), validating the application of free-energy perturbation for this small system as a proof-of-principle test case, albeit at a high computational cost. In one case (conformation 6, at $\Delta A_{FEP} \approx -10$ kcal/mol) the results did not agree. Careful examination of the morph path revealed that one of the backbone angles passes through a highly unfavorable region. The lack of conformational overlap of critical intermediate structures leads to inaccuracies in the averages obtained from the free-energy perturbation simulation. In fact this simulation is the only one for which the results of forward and backward perturbation calculations differ significantly (by ∼1 kcal/mol) whereas they are in near perfect agreement for the remaining simulations. The confinement method, which does not rely on a defined path, does not suffer from this problem. This illustrates the possible dangers when using unrealistic integration paths to calculate free-energy differences by FEP.

Figure 4 shows the excellent convergence of the calculated relative free energies with increasing final restraint constant k_f, indicating that the harmonic approximation appears to be valid for $k_f > 5$ kcal/mol/A^2.

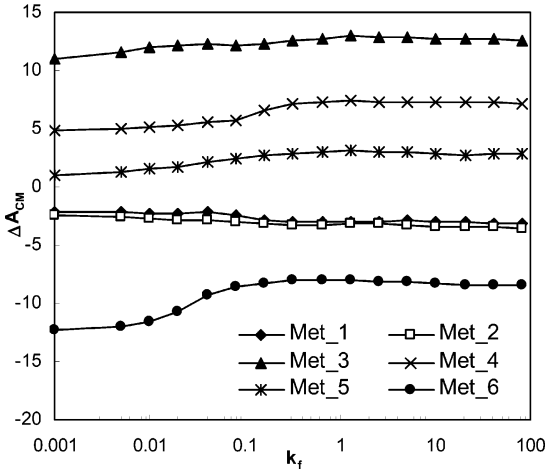

Figure 4. Free-energy differences between conformation Met_0 and the remaining six conformations are shown as a function of varying restraint strength k_f. As expected, the computed free energies converge to a stable value at higher confinement strength.

3.2. Results for U(1−17)T9D. The calculated internal and free energies for the 14 conformations as well as their entropies are shown in Table 2. All values are given relative to conformation Ubi_14, which is in a fully extended state and thus is expected to provide a good reference state in which the side-chain conformational entropy is maximized. Thus entropy changes are negative for all the folded or partially folded conformations, but substantial variations between conformations are evident, demonstrating that side-chain entropy is a major contributing factor to the overall stability of conformations. As expected, the relative entropy and internal energies are inversely correlated (correlation coefficient −0.7). The entropy of the fully α-helical conformation Ubi_13 is comparable to the entropies of the structures forming β-hairpin structures (Ubi_01−Ubi_04). In contrast, those structures that form more globular, compact states (e.g., Ubi_07 and Ubi_11), in which side chains are more directly involved in intramolecular contacts, have considerably lower entropies. The lowest-free-energy structures (including the native conformations) achieve their overall stability by the best trade off between low internal energies while minimizing the conformational entropy loss incurred by forming favorable nonbonded interactions. The latter are mainly realized by backbone−backbone interactions while the side chains extend mostly into the solvent. Despite well-known deficiencies of the implicit solvent model used here and the neglect of backbone entropy differences, it is encouraging that one of the structures with the native conformation as determined by NMR[51] has the lowest free energy in our calculations. It is interesting to note that the native structures were not the lowest in free energy without introduction of the ad hoc solvent−solute van der Waals correction. Instead, the more compact, globular structures were favored, scoring highly favorable van der Waals energies despite the side-chain entropy losses incurred. This highlights the deficiencies in the nonpolar part of the GB/SA model that have been noted before.[55,56]

The fact that the entropy in the fully helical conformation is comparable to that of the β-hairpins is interesting in light of the suggestion that entropic differences on helix formation rationalize helix formation propensities. Creamer and Rose[27] calculated entropy losses for different hydrophobic amino acids and showed a good correlation with experimental data of helix−coil free-energy differences from host−guest studies. While entropy losses on helix formation clearly contribute to the secondary structure propensity of different sequences, entropy losses on β-sheet formation are equally important but much less well studied. Differences in side-chain entropy may be implicated in determining not only propensities to form regular secondary structure but may also determine the sequence-dependent tendency to form other structural elements such as diverging turns, which may be crucial in defining the overall topology of a protein.[31,57] In such conformations, primary interactions occur mainly through side chain−side chain or side chain−backbone contacts, and thus the balance of entropy losses and enthalpy gains will be central to the conformational preferences of such key regions.

Although thermodynamic variables are not strictly separable into component energies due to multibody effects, it can nevertheless be informative to do so approximately. The method presented in this paper allows straightforward dissection of the free energies and internal energies as the thermodynamic integration can be performed on groups of atoms separately. Equally, interaction energies can be approximately separated onto the atoms involved and averages taken separately. For the purpose of free-energy decomposition we ignore the contributions from the normal-mode calculation on the final restrained state ($k = 81.920$ kcal/mol/Å^2) since it is not directly separable into residue components and its contribution to free-energy differences is relatively small (<1.0 kcal/mol). It must be stressed that this simple procedure ignores multibody effects and other correlations and hence is an approximation. The resulting residue entropies of Ubi_02 and Ubi_11 (relative to those of conformation Ubi_14) are projected graphically onto the parent structures in Figure 5.

In the β-hairpin conformation, all side chains are fully solvent exposed and highly mobile, especially in the turn region (e.g., Leu7) and on the convex side of the sheet, Phe3, Thr11, and Thr13. In contrast Glu15 incurs the highest entropy loss due to a strong interaction with the amino terminus. The most striking feature in the compact conformation Ubi_11 is Lys5, which is tightly packed in a surface groove making very favorable interactions with the carboxy terminus, backbone carbonyl groups, and several threonine side chains. This however freezes out the entire side chain, losing almost 5 kcal/mol of entropic stabilization. Other residues involved in this central cluster also contribute to the overall low entropy of this conformation. This is further illustrated in Figure 6 where the average occupancy of different side-chain angles is plotted as a function of simulation time for Ubi_11. The residues interacting in the small core of the structure (Lys5, Thr6, Thr11, and Thr13) are locked in a single rotameric state. Surface residues Gln1, Leu7, Ile12, and Glu15 occasionally sample other conformations but show strong preference to one of the rotamers due to steric clashes or favorable interactions with other parts of the molecule. The remainder of the residues are conformationally free and appear to sample their accessible rotamers during the simulations, showing reasonably good convergence in most cases. However, rotamer distributions of some residues (e.g., Val4) have not fully converged, and longer simulations would be necessary to obtain more precise results.

3.3. Comparison to Normal-Mode and Quasi-Harmonic Analysis.. Previously, side-chain entropies of protein systems have been calculated using both normal-mode analysis[34] and quasi-harmonic analysis.[58] The former assumes that the energy surface around a minimum can be approximated by a quadratic function and ignores diffusive modes while the latter assumes that the atomic fluctuations measured by the covariance matrix of fluctuations can described by a Gaussian probability distribution. Quasi-harmonic analysis has been tested on various

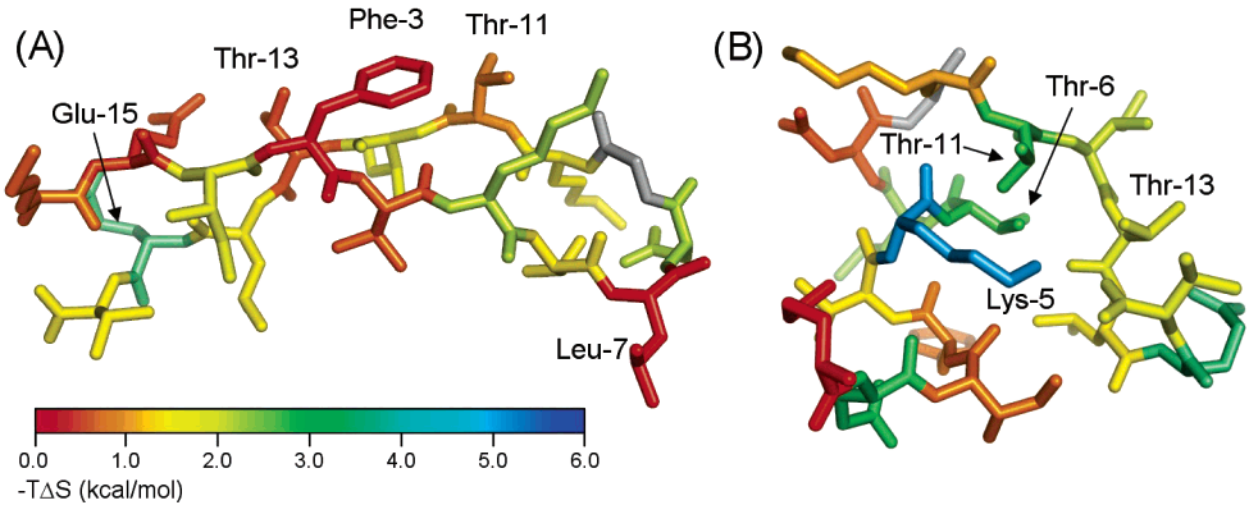

Figure 5. Approximate residue side-chain entropies relative to the extended state Ubi_14 projected graphically onto the respective structures (A) Ubi_02 and (B) Ubi_11. Glycine residues are colored in gray. Entropies are given as $-T\Delta S$ in kcal/mol ($T = 300$K). Figure prepared with Pymol (http://www.pymol.org).

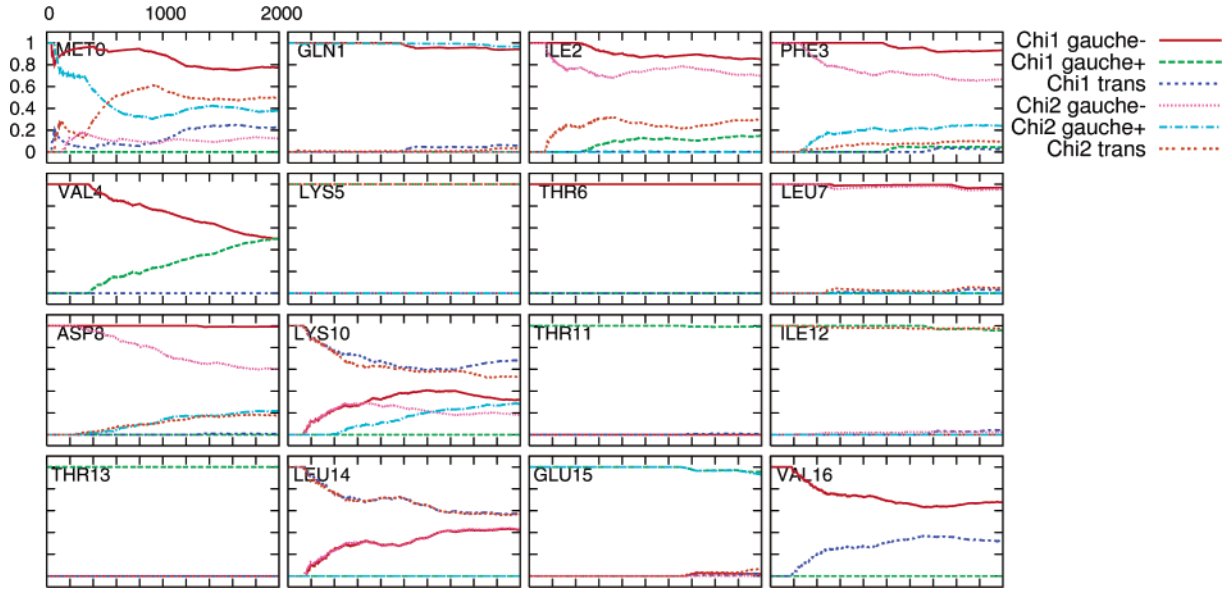

Figure 6. Average side-chain angle distributions (Y-axis showing relative populations of each type of side-chain angle) as a function of simulation length (in ps). Distributions are shown for each residue of conformation Ubi_11 at restraint constant $k = 0.001$.

systems[40,41] and has been shown to produce large errors when the underlying assumptions break down.

To investigate the validity of the above assumptions when applied to peptides, we calculated entropies for our model system with both methods. The normal-mode analysis was performed on the well-minimized starting conformations (as described in the Methods section). Quasi-harmonic analysis[36,59] was performed for each structure on a 10 ns Langevin dynamics simulation by calculating the covariance matrix σ with elements

$$\sigma_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle \tag{11}$$

No fitting was performed since the backbones of the structures were restrained to their initial coordinate positions already. The entropy S_{QH} was then calculated from the eigen-frequencies ω_i of the mass-weighted covariance matrix

$$\det\left(\mathbf{M}^{1/2}\sigma\mathbf{M}^{1/2} - \frac{kT}{\omega^2}\mathbf{1}\right) = 0 \tag{12}$$

and the formula for the harmonic oscillator

$$S_{QH} = R\sum_i \frac{\hbar\omega_i/kT}{\exp(\hbar\omega_i/kT) - 1} - \ln[1 - \exp(-\hbar\omega_i/kT)] \tag{13}$$

Since normal-mode analysis is based on extrapolation of the second derivative matrix of the energy of the local minimum, one would expect normal-mode analysis to underestimate the entropy due to the lack of treatment of diffusive degrees of freedom, such as alternative conformations available to each side chain. Equally, one would expect quasi-harmonic analysis to overestimate the entropy because the Gaussian approximation will fit an inappropriately wide distribution on the distribution of states arising from several discrete minima. Entropies calculated with the two methods are plotted against entropies from the confinement method in Figure 7 (all relative to conformation Ubi_14) and show that the deviations are indeed as expected. The linear fits have gradients of 2.6 and 0.79,

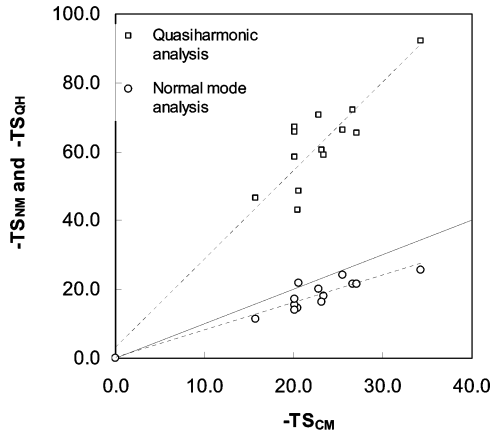

Figure 7. Comparison of entropies calculated using quasi-harmonic analysis ($-TS_{QH}$) and from normal-mode analysis ($-TS_{NM}$) against those calculated from the confinement method ($-TS_{CM}$). All values are relative to conformation Ubi_14 and given in kcal/mol.

respectively. Our results are consistent with other studies investigating the use of the quasi-harmonic approximation on rotationally and torsionally mobile molecules.[41,42]

Although we utilize an implicit solvent model with well-known deficiencies, the results demonstrate, in principle, the importance of including the entropy of ensembles of conformations when assessing the relative stability of protein conformations. Typically, single-point energies or internal energies are used for the purpose of ranking structures. Energies from single conformations can only be used when the respective force field is very smooth and essentially seeks to empirically approximate free energies of ensembles by averaging implicitly over some of the degrees of freedom.[60−63] The hardness of a typical all-atom force field however makes a single-point energy, even of a minimum,[64] a poor measure of the stability of the conformation, because miniscule changes in structure can change the potential energy dramatically. Some studies have used a short molecular dynamics trajectory to obtain a more stable, averaged potential energy to assess the stability of large decoy sets.[65] This approach assumes that the relative differences in conformational entropy between compact protein structures are negligible when searching exclusively for the most stable conformation.[66,67] Both studies however ignore the configurational side-chain entropy, by including only entropies calculated from C_α fluctuations using quasi-harmonic analysis or vibrational entropies from normal-mode analysis, respectively. While the native state of larger proteins may have a sufficiently large internal energy gap to the next lowest energy structure, allowing entropic differences to be ignored, this does not appear to be a valid assumption for small proteins/peptides where the relative stabilities vary only by a few kcal/mol. Thus to investigate the physicochemical basis for local structure propensity it is essential to include all entropic differences. Empirical models to account for configurational side-chain entropy based on rotamer counting[68,69] have been proposed in the past but are arbitrary with respect to choice of rotamer library and assessment of population of different rotameric states and ignore vibrational entropy changes. The method presented here, although not suitable for rapid scoring of conformations, provides a rigorous approach to side-chain entropy and will allow parametrization of new, fast empirical functions.

4. Conclusions

A novel, path-independent method for the calculation of free-energy differences that requires simulation only at the "end" states of interest has been presented, tested against an established approach on a small model system, and then applied to the rigorous calculation of side-chain entropies of various conformations of a stable β-hairpin. The results demonstrate the importance of side-chain entropy in structural preferences of peptides. The method will be applicable to a wide range of free-energy problems and should prove particularly useful where the definition of a suitable integration path is problematic or where the path is exceedingly long, thus preventing efficient use of classical approaches. For example ligand-binding reactions in which large conformational changes occur upon binding present a formidable challenge to free-energy methods, although some steps have been made toward that goal.[70] Such processes are of central interest in biochemistry since conformational changes play a major role in function and regulation of virtually all cellular processes. The method presented in this paper may provide a way to approach these problems and to extend the scope of free-energy calculations.

Acknowledgment. M.D.T. gratefully acknowledges funding support from the Department of Biochemistry, University of Bristol, and the authors thank Rhiju Das for helpful comments on the manuscript.

References and Notes

(1) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.
(2) Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(3) van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Helv. Chim. Acta* **2002**, *85*, 3113.
(4) Chipot, C.; Pearlman, D. A. *Mol. Simul.* **2002**, *28*, 1.
(5) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(6) Miller, J. L.; Kollman, P. A. *J. Phys. Chem. A* **1996**, *100*, 8587.
(7) Henchman, R. H.; Essex, J. W. *J. Comput. Chem.* **1999**, *20*, 499.
(8) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
(9) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(10) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(11) Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 307.
(12) Oostenbrink, C.; van Gunsteren, W. F. *Proteins* **2004**, *54*, 237.
(13) Patey, G. N.; Valleau, J. P. *J. Chem. Phys.* **1975**, *63*, 2334.
(14) Torrie, G. M.; Valleau, J. P. *J. Comput. Chem.* **1977**, *23*, 187.
(15) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169.
(16) Boczkot, E. M.; Brooks, C. L., III. *J. Phys. Chem.* **1993**, *97*, 4509.
(17) Chipot, C.; Hénin, J. *J. Chem. Phys.* **2005**, *123*, 244906.
(18) Lee, M. S.; Olson, M. A. *Biophys. J.* **2006**, *90*, 864.
(19) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568.
(20) Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825.
(21) Hermans, J.; Shankar, S. *Isr. J. Chem. Phys.* **1986**, *27*, 225−227.
(22) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys J.* **1996**, *71*, 670.
(23) Zhang, L.; Hermans, J. *Proteins* **1996**, *24*, 433.
(24) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707.
(25) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535.
(26) Strajbl, M.; Sham, Y. Y.; Villa, J.; Chu, Z.-T.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 4578.
(27) Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5937.
(28) Richardson, J. M.; Lopez, M. M.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1413.
(29) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068.
(30) Brooks, C. L., III. *Acc. Chem. Res.* **2002**, *35*, 447.
(31) Gnanakaran, S.; Garcia, A. E. *Biophys. J.* **2003**, *84*, 1548.
(32) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168.
(33) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766.
(34) Ma, B.; Tsai, C.; Nussinov, R. *Biophys. J.* **2000**, *79*, 2739.
(35) Karplus, M.; Kushick, J. *Macromolecules* **1981**, *14*, 325.

(36) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370.
(37) Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894.
(38) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *119*, 9947.
(39) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
(40) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809.
(41) Carlsson, J.; Aqvist, J. A. *J. Phys. Chem. B* **2005**, *109*, 6448.
(42) Chang, C.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, 1017.
(43) Villa, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z.-T.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11899.
(44) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(45) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(46) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
(47) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(48) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222.
(49) Paterlini, M. G.; Ferguson, D. M. *Chem. Phys.* **1998**, *236*, 243.
(50) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(51) Zerella, R.; Chen, P. Y.; Evans, P. A.; Raine, A.; Williams, D. H. *Protein Sci.* **2000**, *9*, 2142.
(52) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(53) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(54) Zhang, W.; Hou, T.; Qiao, X.; Xu, X. *J. Phys. Chem. B* **2003**, *107*, 9071.
(55) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 9523.
(56) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479.
(57) Qian, Y.; Bystroff, C.; Rajagopal, P.; Klevit, R. E.; Baker, D. *J. Mol. Biol.* **1998**, *283*, 293.
(58) Schafer, H.; Smith, L. J.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2002**, *46*, 215.
(59) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
(60) Czaplewski, C.; Oldziej, S.; Liwo, A.; Scheraga, H. A. *Protein. Eng., Des. Sel.* **2004**, *17*, 29.
(61) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421.
(62) Gibbs, N.; Clarke, A. R.; Sessions, R. B. *Proteins* **2001**, *43*, 186.
(63) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534.
(64) Hsieh, M.; Luo, R. *Proteins* **2004**, *56*, 475.
(65) Lee, M. C.; Duan, Y. *Proteins* **2004**, *55*, 620.
(66) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins* **1998**, *32*, 399.
(67) Lee, M. R.; Duan, Y.; Kollman, P. A. *Proteins* **2000**, *39*, 309.
(68) Abagyan, A. R. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Klewer Academic Publishers: Dordrecht, 1997; Vol. 3; p 363.
(69) Pickett, S. D.; Sternberg, M. J. E. *J. Mol. Biol.* **1995**, *231*, 825.
(70) Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M. *J. Mol. Biol.* **2005**, *353*, 196.