Calculation of the "absolute" free energy of a β-hairpin in an all-atom force field

Cite as: J. Chem. Phys. **139**, 054102 (2013); https://doi.org/10.1063/1.4817195
Submitted: 21 March 2013 • Accepted: 17 July 2013 • Published Online: 01 August 2013

Konstantin V. Klenin and Wolfgang Wenzel

Calculation of the "absolute" free energy of a β-hairpin in an all-atom force field

Konstantin V. Klenin[1,a)] and Wolfgang Wenzel[2]
[1]*Steinbuch Centre for Computing, Karlsruhe Institute of Technology, P.O. Box 3640, D-76021 Karlsruhe, Germany*
[2]*Institute of Nanotechnology, Karlsruhe Institute of Technology, P.O. Box 3640, D-76021 Karlsruhe, Germany*

(Received 21 March 2013; accepted 17 July 2013; published online 1 August 2013)

We propose a new approach to calculate the conformational free energy of a macromolecule in a compact stable state in implicit solvent. The free energy is evaluated with respect to an artificial reference system without internal interactions that is confined to a small well-defined multidimensional volume of a regular shape occupying approximately the same part of the conformational space as the macromolecule of interest. We present a practical implementation of our method, successfully apply it to a β-hairpin in all-atom representation, verify the results with direct parallel tempering simulations, and discuss the possibilities of further improvements. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4817195]

I. INTRODUCTION

The conformational free energy of a certain state of a macromolecular system is an important measure of its stability. Although it is possible to estimate the free energy from numerical simulations, it is not usually easy. As a rule, such estimation is done in an indirect way: one obtains not the "absolute" free energy, but the free energy difference calculated in comparison with some other reference state or reference system. The reliability of the estimation strongly depends on similarity of the two compared states (systems): the less is the similarity, the more difficult it is to obtain a reliable estimation. (For reviews, see, for example, Refs. 1–4.)

In the present study, we consider the possibility to approach the problem from the other side. We are interested in calculation of the "absolute" conformational free energy F without comparison to other physical states or systems. In principle, it can be done by means of the formula

$$\langle e^{\beta U(\mathbf{x})}\rangle = \frac{1}{Z}\int d\mathbf{x}\, e^{-\beta U(\mathbf{x})}\, e^{\beta U(\mathbf{x})} = \frac{V}{Z}, \qquad (1)$$

where $\mathbf{x}$ is a point in the conformational space, $U(\mathbf{x})$ is the potential energy, $\beta = 1/k_{\mathrm{B}}T$ is the inverse Boltzmann constant multiplied by the inverse temperature, $Z = e^{-\beta F}$ is the partition function, and $V = \int d\mathbf{x}$ is the total volume of the conformational space, which we assume to be finite. The average quantity $\langle e^{\beta U(\mathbf{x})}\rangle$ can be, in principle, found in numerical simulations. But, in practice, the accuracy of results in most cases would be absolutely unsatisfactory. The problem here is not that the regions with the highest values of $e^{\beta U(\mathbf{x})}$ would not be sampled, as this inconvenience can be easily eliminated in some non-canonical sampling procedure. The real problem is that one has to sample the whole huge multidimensional conformational space, making the simulations unfeasible.

However, the situation might dramatically improve if the system, in the state of interest, occupies only a small region of the total conformational space. One can confine this region (or its representative part) by a boundary with a regular shape, such as a multidimensional box or ellipsoid, for which the volume V can be easily defined. In this case, the numerical evaluation of the ratio V/Z might become quite feasible, either by direct computation of $\langle e^{\beta U(\mathbf{x})}\rangle$ or by some more elaborate technique.

Here, we report our first implementation of this approach and show that, for certain systems, it can be quite efficient.

II. THE METHOD OF THE "ABSOLUTE" FREE ENERGY COMPUTATION

We consider a macromolecule as a classical system with the potential energy $U(\mathbf{x})$, where $\mathbf{x}$ is a point in the D-dimensional space of internal degrees of freedom measured in some dimensionless units. The solvent is taken into account implicitly. We assume that the system is in a compact stable state, such as the native state of a protein, so that conformational fluctuations are relatively small. Numerical simulation methods (e.g., Monte Carlo) applied to this system produce a set of structures that occupy only a very small part of the total conformational space. Given these structures, we can construct a small D-dimensional ellipsoid that comprises a noticeable fraction (say, ~10%–50%) of the total set. After that, we define another system in such a way that its potential energy inside the ellipsoid is equal to zero, but all the conformations outside it are forbidden. The new system is quite similar, in some respect, to the original one. Its free energy, however, can be calculated analytically, as its partition function is just equal to the volume of the ellipsoid. Taking advantage of the similarity of the two systems, one can relatively easily calculate the free energy difference between them by known methods and, thus, evaluate the free energy of the original system. Below, we describe this procedure in more detail.

For the construction of the ellipsoid we use a technique that is similar to the principal component analysis (PCA).[5]

a)Author to whom correspondence should be address. Electronic mail: konstantin.klenin@kit.edu.

Let us consider N conformations obtained in a simulation run of $(N-1)$ steps as the $D \times N$ matrix $\mathbf{X} = \|x_{dn}\|$, where $d = 1, \ldots, D$ and $n = 1, \ldots, N$, so that the nth column of this matrix represents the conformation $\mathbf{x}_n$ after $(n-1)$ steps. Let $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$ be the average conformation and let $\tilde{\mathbf{X}} = \|x_{dn} - \bar{x}_d\|$ be the matrix of all conformations shifted by the vector $\bar{\mathbf{x}}$. The covariance matrix is defined as $\mathbf{C_X} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{T}}/(N-1)$. We denote its eigenvalues by $\lambda_1, \ldots, \lambda_D$, its normalized eigenvectors by $\mathbf{e}_1, \ldots, \mathbf{e}_D$, and the matrix composed of the normalized eigenvectors (taken as columns) by $\mathbf{E} = \|\mathbf{e}_1 \ldots \mathbf{e}_D\|$. We define the new set of degrees of freedom by the transformation $\mathbf{y} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}})$, where $\mathbf{\Lambda}^{-1/2}$ is the diagonal matrix with the diagonal elements $\lambda_1^{-1/2}, \ldots, \lambda_D^{-1/2}$. As can be verified by straightforward computations, the average value of $\mathbf{y}_n$ is equal to zero and the covariance matrix $\mathbf{C_Y} = \mathbf{Y}\mathbf{Y}^{\mathrm{T}}/(N-1)$ for the transformed dataset $\mathbf{Y} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^{\mathrm{T}}\tilde{\mathbf{X}}$ is equal to the unit matrix. In the space of the $\mathbf{y}$-vectors, we draw now a D-dimensional sphere with the center at the origin $\mathbf{y} = \mathbf{0}$ and the radius R, such that a certain non-zero fraction p of all the data points $\mathbf{y}_n$ lie within this sphere. (In principle, the value of p can be chosen arbitrary. However, this choice will affect the accuracy of the final result. The range $p \sim 0.1$–0.5 seems to be a reasonable suggestion.) The volume of a multidimensional sphere is known to be $V_D(R) = \pi^{D/2}R^D/\,\Gamma(1 + D/2)$, where $\Gamma(\ldots)$ is the gamma-function.[6] In the $\mathbf{x}$-space, the corresponding ellipsoid has the volume

$$V_0 = V_D(R)\prod_{d=1}^{D}\lambda_d^{1/2}. \tag{2}$$

It should be noted that the dimensionality of the ellipsoid remains the same as that of the whole conformational space, despite the use of the PCA, which is usually associated with reduction of degrees of freedom.

As a next step, we can define a set of $(K+1)$ auxiliary systems confined to the interior of the ellipsoid with the energy functions $U_k(\mathbf{x}) = \alpha_k U(\mathbf{x})$, $k = 0, 1, \ldots, K$, where the constants α_k satisfy the inequality $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_K = 1$. The exact values of K and the constants α_k should be chosen in such a way that the free energy differences between the neighboring systems can be reliably evaluated without the need for any other intermediate auxiliary systems. In this study, we calculate these free energy differences in the following way.

Consider a volume v in the conformational space that can be well sampled by the simulations of both the neighboring systems defined by the energies $U_k(\mathbf{x})$ and $U_{k+1}(\mathbf{x})$. Let $Z_k^{(v)}$ and $Z_{k+1}^{(v)}$ be the partition functions of the corresponding subsystems constrained to the given volume. Their ratio can be calculated by the free energy perturbation method:[7]

$$\frac{Z_{k+1}^{(v)}}{Z_k^{(v)}} = \frac{\int\limits_v d\mathbf{x}\exp(-\beta U_k)\exp[-\beta(U_{k+1} - U_k)]}{\int\limits_v d\mathbf{x}\exp(-\beta U_k)} = \langle\exp[-\beta(U_{k+1} - U_k)]\rangle_k^{(v)}, \tag{3}$$

where the notation $\langle\ldots\rangle_k^{(v)}$ stands for averaging over the equilibrium ensemble of the kth subsystem. In terms of the simulation data, Eq. (3) means

$$\frac{Z_{k+1}^{(v)}}{Z_k^{(v)}} = \frac{1}{N_k^{(v)}}\sum_{\mathbf{x}_n^{(k)} \in v}\exp\left[-\beta(\alpha_{k+1} - \alpha_k)U\left(\mathbf{x}_n^{(k)}\right)\right], \tag{4}$$

where the summation is carried out over the conformations $\mathbf{x}_n^{(k)}$ of the kth system within the volume v, with $N_k^{(v)}$ being the total number of such conformations.

In this study, however, we use a more symmetrical expression for the ratio of the partition functions. Introducing an intermediate system with the energy $U_{k+1/2}(\mathbf{x}) = \frac{1}{2}[U_k(\mathbf{x}) + U_{k+1}(\mathbf{x})]$, we have

$$\frac{Z_{k+1}^{(v)}}{Z_k^{(v)}} = \frac{Z_{k+1}^{(v)}}{Z_{k+1/2}^{(v)}}\frac{Z_{k+1/2}^{(v)}}{Z_k^{(v)}} = \frac{\langle\exp[-\beta(U_{k+1/2} - U_k)]\rangle_k^{(v)}}{\langle\exp[-\beta(U_{k+1/2} - U_{k+1})]\rangle_{k+1}^{(v)}}. \tag{5}$$

The partition function of the kth system can be expressed as $Z_k = (N_k/N_k^{(v)})Z_k^{(v)}$, where N_k is the total number of conformations in the simulation dataset. Hence, for the ratio of the partition functions of the neighboring systems, after performing the averaging as in Eq. (4), we get from Eq. (5)

$$\frac{Z_{k+1}}{Z_k} = \frac{N_{k+1}\sum\limits_{\mathbf{x}_n^{(k)} \in v}\exp\left[-\frac{1}{2}\beta(\alpha_{k+1} - \alpha_k)U\left(\mathbf{x}_n^{(k)}\right)\right]}{N_k\sum\limits_{\mathbf{x}_n^{(k+1)} \in v}\exp\left[-\frac{1}{2}\beta(\alpha_k - \alpha_{k+1})U\left(\mathbf{x}_n^{(k+1)}\right)\right]}. \tag{6}$$

Note that the both summations are still carried out only for the conformations belonging to the volume v. Practically, this volume can be conveniently defined by the condition that, for $\mathbf{x} \in v$, the value of the function $U(\mathbf{x})$ belongs to a certain interval, $U_{k,k+1}^{\min} < U(\mathbf{x}) < U_{k,k+1}^{\max}$, chosen in such a way that the density function for the probability distribution of $U(\mathbf{x})$ can be reliably estimated everywhere within it for both the neighboring systems.

Combining all together, we can now write down the final expression for the free energy of the original system:

$$F = -k_{\mathrm{B}}T\ln\left(\frac{1}{p}\frac{Z_K}{Z_{K-1}}\frac{Z_{K-1}}{Z_{K-2}}\ldots\frac{Z_1}{Z_0}V_0\right), \tag{7}$$

where the volume V_0 of the confining ellipsoid is defined by Eq. (2) and the ratios Z_{k+1}/Z_k are the quantities obtained from simulations according to Eq. (6).

III. IMPLEMENTATION FOR A β-HAIRPIN

A. Simulation details

For the purpose of illustration we applied our approach to Trp zipper 1, a small polypeptide that forms a β-hairpin at room temperature.[8] The Monte Carlo (MC) simulations were performed using SIMONA program[9] and the all-atom force field PFF02 with implicit solvent[10] that was shown to reasonably reproduce the thermodynamic properties of Trp zipper 1.[11] In order to prevent too large values of $U(\mathbf{x})$ for the auxiliary system with $\alpha_0 = 0$, we modified the force field in such a way that the absolute values of the Lennard-Jones and

electrostatic energies of interaction between any two atoms were limited to 10 kcal/mol. All the lengths and angles of the covalent bonds were fixed, and the only degrees of freedom were the dihedral angles of the backbone and side chains. The total number of degrees of freedom was $D = 48$. The following two types of MC moves were used: (1) the random change of a dihedral angle in the interval $(-15^\circ, 15^\circ)$ and (2) the local perturbation of the backbone within four consecutive residues based on the exact loop closure algorithm.[12]

B. Folded ensemble

The calculation of the free energy of the native β-hairpin was done for the temperature $T = 160$ K, which is quite below the folding transition mid-point that was estimated to be ~230 K. (By the temperature we mean just the formal parameter used in our simulations with the given force field without any relation to realistic values.) The set of systems with different α_k values was simulated in a single parallel tempering run, since, in the frame of the MC procedure, the system with the energy $\alpha_k U(\mathbf{x})$ at the temperature T is equivalent to the system with the energy $U(\mathbf{x})$ at the temperature $T_k = T/\alpha_k$. A conformation was considered to belong to the native state if its root-mean-square deviation (RMSD) from the first structure deposited in the Protein Data Bank (PDB) under the code 1LE0 did not exceed 4.3 Å.

First of all, we performed parallel tempering simulation of the system constrained to the native state. (Although the parallel tempering is formally not required at this stage, we still used this technique to provide better quality of results.) From the total trajectory of 5×10^7 MC steps we registered at equal intervals 10 000 conformations corresponding to the temperature $T = 160$ K. These conformations were available in terms of $D = 48$ dihedral angles x_d, which had to be put in an interval $(x_d^{\min},\ x_d^{\min} + 2\pi)$ optimized for the further PCA-like procedure of the ellipsoid construction. For each angle x_d, we built a histogram of $m = 100$ bins and defined the $x_d^{\min}$ value as the mid-point of the longest series of consecutive bins for which the occupancy was equal or below the threshold of $h = 0.36\%$. If the occupancies of all the bins were above this threshold, then $x_d^{\min}$ was taken as the mid-point of the lowest bin. (We used this simplified procedure with the given values of m and h for historical reasons. A stricter approach to the PCA in the space of dihedral angles can be found, for example, in Ref. 13.) After the transformation to the **y**-space (described above), we calculated the average distance of the **y**-points to the origin, which turned to be 6.6086, and took this value as the radius R of the constraining sphere. The fraction of the points lying within it was $p = 0.646$. Note that the radius R is quite close to the ideal theoretical value of $\sqrt{D} = 6.9282$ calculated under the assumption that each component of the **y**-vector has a Gaussian distribution.

The second parallel tempering run was performed for the system that, in the **y**-space, was confined to the given sphere and, in the **x**-space, to the corresponding ellipsoid. Although the constraint for the RMSD was now formally discarded, it was actually fulfilled for all the conformations (otherwise one would have to introduce an additional correction factor into Eq. (7)). The MC trajectory consisted of 5×10^7 steps. The temperature range was from 160 to 10^{20} K, the latter value serving as an approximation to infinity. After this run we were able, in the frame of our model, to calculate the free energy F_{nat} of the native state according to Eq. (7).

C. Unfolded ensemble

The polypeptide chain of Trp zipper 1 turned out to be small enough to allow calculation of the free energy F_{unf} of the unfolded state in a similar way, though without construction of the ellipsoid. The third parallel tempering run, of 5×10^7 steps, was carried out for the system with forbidden native state (RMSD > 4.3 Å) in the temperature range from 160 to 2500 K. In addition, three non-parallel runs of the same length were performed with the infinite temperature: the first one with the energies constrained to the values below the threshold of $h'' = 80$ kcal/mol, the second one with the threshold of $h' = 200$ kcal/mol, and the last one without any threshold. The RMSD constraint was not necessary this time, as the occurrence of the native state was highly improbable. The partition function Z_0 in the latter case is obviously equal to $(2\pi)^D$ with a very good accuracy, as the fraction of the conformational space corresponding to the native state (which should be excluded) is negligible. The partition function of the system with the threshold h' was calculated from the relation $Z_0' = p' Z_0$, where p' is the fraction of the conformations with the energy below h' in the unconstrained ensemble. In a similar way, the partition function of the system with the threshold h'' was obtained as $Z_0'' = p'' Z_0'$, where p'' is the fraction of the conformations lying below the threshold h'' in the ensemble constrained by the threshold h'. The ratio Z_1/Z_0'', where Z_1 is the partition function of the replica with the highest temperature in the parallel tempering run, was determined in the standard way according to Eq. (6), as the corresponding ensembles were sufficiently overlapping. The free energy of the unfolded state was found as

$$F_{\text{unf}} = -k_{\text{B}} T \ln\left(\frac{Z_K}{Z_{K-1}}\frac{Z_{K-1}}{Z_{K-2}}\cdots\frac{Z_1}{Z_0''}p''p'(2\pi)^D\right), \qquad (8)$$

where Z_k is the partition function of the kth replica, $k = 1, \ldots, K$, so that $Z_K = Z_{\text{unf}}$ corresponds to the lowest temperature $T = 160$ K.

At last, a control parallel tempering run, of 1.5×10^8 steps, without any constraints was performed in the temperature range from 160 to 350 K, which allowed direct calculation of the folding free energy $\Delta F_{\text{fold}}^{\text{dir}} = k_{\text{B}} T \ln[p_{\text{unf}}/(1 - p_{\text{unf}})]$, where p_{unf} is the fraction of unfolded conformations at $T = 160$ K.

IV. RESULTS

The β-hairpin structure of Trp zipper 1, which was used as a model system in the present study, is shown in Fig. 1. We estimated the free energy of the native state of this polypeptide from the data obtained by MC simulations in the all atom force field PFF02 at the reduced temperature $T = 160$ K, quite below the folding mid-point (~230 K). The dimensionality of this system is $D = 48$, with the dihedral angles being the only degrees of freedom. After the preliminary simulation, we confined the further simulations to a D-dimensional ellipsoid

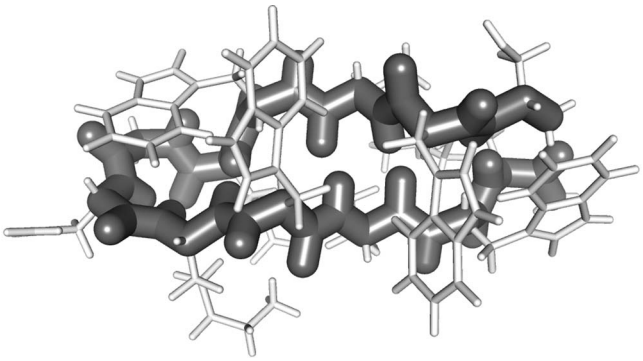

FIG. 1. The native structure of Trp zipper 1 (the first model deposited in the PDB under the code 1LE0).

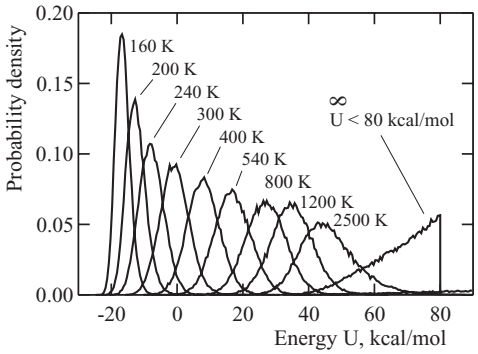

FIG. 3. The probability density functions for the potential energy $U(\mathbf{x})$ at different temperatures T_k for unfolded conformations.

comprising considerable part of the native state ensemble, as described above. The probability distribution functions for the potential energy of the confined system at different temperatures T_k are presented in Fig. 2. Since the regions where these functions can be reliably evaluated are overlapping, one can use Eqs. (2), (6), and (7), with $\alpha_k = T/T_k$, to estimate the free energy F_{nat} of the native state. We found that F_{nat} was equal to 8.16 ± 0.02 kcal/mol. Here and below, the presented numerical values of the free energies are corrected by subtracting the constant term $-k_{\text{B}}T\ln(2\pi)^D$, where $(2\pi)^D$ is the volume of the total conformational space, so that these values are independent from the measurement units used for the degrees of freedom.

The distribution functions of the potential energy for the unfolded state at different temperatures T_k are displayed in Fig. 3. The overlapping between "neighboring" distributions is sufficient for reliable evaluation of the free energy according to Eq. (8). The free energy of the unfolded state at the reduced temperature $T = 160$ K was found to be $F_{\text{unf}} = 9.49 \pm 0.06$ kcal/mol. Consequently, the free energy of folding is equal to $\Delta F_{\text{fold}} = F_{\text{nat}} - F_{\text{unf}} = -1.33 \pm 0.06$ kcal/mol. This value is to be compared with the estimation obtained from the direct parallel tempering simulations of the unconstrained system, which yielded $\Delta F^{\text{dir}}_{\text{fold}} = -1.24 \pm 0.10$ kcal/mol. Taking into account the accuracy of the calculations, the agreement is quite good. Note that the highest accuracy was achieved for the compact native state, for which our approach is best suited.

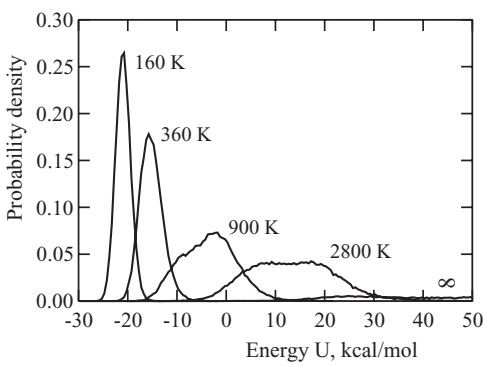

FIG. 2. The probability density functions for the potential energy $U(\mathbf{x})$ at different temperatures T_k for the system confined to the ellipsoid that occupies approximately the same part of the conformational space as the native state.

V. DISCUSSION

The presented method for calculation of the free energies proved to be quite efficient for the short polypeptide Trp zipper 1, which was taken as a model case. In general, however, one can expect that the applicability of our approach would strongly depend on the particularities of the considered system. The more compact is the given state of a system, the more reliable is the method.

In our calculations, the "absolute" free energy of the folded hairpin was estimated with the accuracy of 0.02 kcal/mol. Surprisingly enough, it was also possible to get the "absolute" free energy of the unfolded state, in which case, however, the statistical error was noticeably larger, 0.06 kcal/mol, whereas the computational efforts were approximately the same. As the unfolded state is far from being compact, no confining ellipsoid could be constructed and, hence, the procedure was reduced to the usual umbrella sampling.[14] Our method, in its pure form, is most suitable for the case when the free energies of two compact stable states of the same protein should be compared. This kind of problem may occur, for example, in the course of protein structure prediction, when one has to decide between two or more dissimilar decoys with close energy values.[15]

The present calculation of the folding free energy was performed for the purpose of illustration and verification, because the result could be easily compared with that obtained by the direct simulations by parallel tempering MC. We find it encouraging that, for the given system, our approach yielded higher accuracy than 0.10 kcal/mol achieved in the direct simulations that took approximately equal computational time. The gain in efficiency is particularly noticeable for the compact folded state.

In the present study, we applied canonical weights to protein conformations and kept the energy landscape very ragged, which is quite natural to start with, but obviously not optimal. (The cutoff of 10 kcal/mol used for Lennard-Jones and electrostatic pairwise interactions was far from being enough to eliminate raggedness.) We see further possible improvement of the method in reweighting (or smoothing) the clashing conformations, so that the whole confining ellipsoid can be efficiently sampled in very few simulation runs, which is not unrealistic for a compact subsystem. This question is now under investigation.

In principle, the sampling of the confining ellipsoid can be improved by reduction of its size. This can be achieved, for example, by reiterating the procedure of ellipsoid construction: new full-energy simulations restrained to the first ellipsoid are performed, and the obtained set of conformations is used to build the second ellipsoid that comprises only a small part of this set, and so on. We find, however, that the costs of such reiterations are considerably larger than the resulting gains, at least for the chosen model system. It should be noted that iterative restraining of a system to decreasingly smaller regions in the conformational space is the key procedure of the deactivated morphing,[16] which is another technique of free energy computation. However, in the frame of the deactivated morphing the high costs of this procedure are justified, as they are compensated by the possibility to treat the explicit solvent.

In addition to the dihedral angles considered in this study, other types of degrees of freedom can be used as well. The only requirement is that they should be uniformly distributed when the potential energy is set to zero: $U(\mathbf{x}) \equiv 0$. Thus, instead of the bond angle γ, one should take $\cos\gamma$ and instead of the bond length r, the quantity r^3, or, speaking more strictly, $(r/r_0)^3$, where r_0 is some constant length (for example, 1 Å), as the degrees of freedom are supposed to be dimensionless. In the case when the system consists of two associated molecules, the translational degrees of freedom can be used straightforwardly (divided by r_0), but the set of the Euler angles (φ, θ, ψ) should be substituted by $(\varphi, \cos\theta, \psi)$.

The presented method is compatible with the Brownian dynamics and, in principle, with the molecular dynamics (with a suitable thermostat). As the Brownian dynamics maintains the detailed balance, it can be formally considered as a special case of MC, and, hence, the rejection of the moves leading out from the confining ellipsoid should be performed according to the usual MC rules: after rejecting a move, the system returns to the previous state, which is counted in the generated ensemble for another time. In molecular dynamics, keeping the system inside the ellipsoid is a more delicate matter as it additionally may involve reinitialization of velocities after the rejected move.

One should mention that the force field PFF02 used in this study was originally designed for prediction of protein structures. It was shown that its global minimum is capable to reproduce the experiment native structure of proteins.[10] However, any scaling factor can be applied simultaneously to all terms of this force field without affecting this capability. For that reason, the simulation temperatures are given here as formal parameters. For example, the denaturation mid-point of Trp zipper 1 was estimated to be ∼230 K, whereas the experimental value[8] is 323 K, which yields the correction coefficient of 1.4. Thus, for the PFF02, the room temperature corresponds to 213 K. The folding probability at this temperature extracted from our simulations is 0.70 (data not shown), being in a good agreement with the experimental value of 0.73 reported by Cochran *et al.*[8] The free energy barrier between the folded and unfolded states is not particularly high, so that one can directly observe folding-unfolding events in ordinary MC simulations. In order to avoid this trivial situation, we, for the purpose of illustration, took a lower temperature, 160 K. Extrapolation of experimental data to this temperature using Eq. (4) of Ref. 17 gives the folding probability of 0.22, noticeably smaller than the value 0.98 corresponding to our free energy estimations. This disagreement is, however, quite expectable, as the force field parameters do not depend on the temperature, so that the effect of cold denaturation cannot be reproduced.

The only essential limitation of our method is the requirement of implicit solvent. Although the force fields with implicit solvent are widely used in simulations of proteins and other macromolecules,[18] they are not capable to capture some fine effects of solute-solvent interactions. This limitation, however, will be removed as soon as we have an efficient procedure to estimate the free energy difference between a system with implicit solvent and its counterpart with explicit solvent. This problem, however, does not seem to be difficult.

A typical force field with explicit solvent contains the following terms:

$$U^{\mathrm{expl}}(\mathbf{x}, \mathbf{s}) = U_{\mathrm{p}}(\mathbf{x}) + U^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x}, \mathbf{s}) + U^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x}, \mathbf{s}) + U_{\mathrm{s}}(\mathbf{s}), \qquad (9)$$

where $\mathbf{x}$ represents, as before, a point in the conformational space of a protein, $\mathbf{s}$ denotes the coordinates of the solvent molecules, $U_{\mathrm{p}}(\mathbf{x})$ and $U_{\mathrm{s}}(\mathbf{s})$ are the energies of internal interactions within the protein and the solvent, respectively, whereas the external interactions between the protein and the solvent are represented by the electrostatic, $U^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x}, \mathbf{s})$, and Lennard-Jones, $U^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x}, \mathbf{s})$, potentials. A typical force field with implicit solvent has the form

$$U^{\mathrm{impl}}(\mathbf{x}) = U_{\mathrm{p}}(\mathbf{x}) + F^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x}) + F^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x}), \qquad (10)$$

where $U_{\mathrm{p}}(\mathbf{x})$ is the same term as in Eq. (9) and the terms $F^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x})$ and $F^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x})$ approximate the free energies of electrostatic and surface interactions with the solvent, respectively. Consider the "intermediate" system given by the conformational energy:

$$U^{\mathrm{interm}}(\mathbf{x}, \mathbf{s}) = U_{\mathrm{p}}(\mathbf{x}) + F^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x}) + U^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x}, \mathbf{s}) + U_{\mathrm{s}}(\mathbf{s}). \qquad (11)$$

Formally, it treats the solvent explicitly; however, the external electrostatics is substituted by the implicit term $F^{\mathrm{E}}_{\mathrm{ps}}(\mathbf{x})$. In this system, the explicit protein-solvent interactions are so simplified that they can be approximated with high accuracy by the implicit solvent model described by Eq. (10), where the term $F^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x})$ is just proportional to the surface area $S(\mathbf{x})$ of the protein: $F^{\mathrm{LJ}}_{\mathrm{ps}}(\mathbf{x}) = \sigma S(\mathbf{x})$, with the coefficient σ being a macroscopic property of the solvent. On the other hand, the "intermediate" system is quite similar to the system with explicit solvent (Eq. (9)), so that the corresponding free energy difference ΔF can be estimated by the standard techniques. The quantity ΔF is to be used as a correction to the "absolute" free energy F of the system with implicit solvent (Eq. (10)), which is available by our method.

VI. CONCLUSIONS

We considered the problem of free energy estimation for a macromolecule in the implicit solvent approximation. If the molecule is in a compact stable state with relatively small structural fluctuations, one can take the advantage of

this compactness to calculate the "absolute" free energy by the new method proposed in this study. The epithet "absolute" means that the free energy evaluation is done not in comparison with some other physical state, but rather in comparison with an artificial reference system which is simple enough to allow analytical calculation of its free energy. The simplicity of the reference system is achieved by switching off all internal interactions and by confining its conformations to a small multidimensional ellipsoid in the same conformational space region that is occupied by the macromolecule of interest.

The new method is easy to implement. We demonstrated its efficiency and reliability for a model system of 48 degrees of freedom and discussed the possibilities of its further improvement. We believe that the new approach will be a useful contribution to the existing repertoire of the efficient methods for calculation of the free energy.

ACKNOWLEDGMENTS

We would like to thank Dr. Ivan Kondov for fruitful discussions.

[1]C. D. Christ, A. E. Mark, and W. F. van Gunsteren, J. Comput. Chem. **31**, 1569 (2010).
[2]A. Pohorille, C. Jarzynski, and C. Chipot, J. Phys. Chem. B **114**, 10235 (2010).
[3]M. R. Shirts and D. L. Mobley, Methods Mol. Biol. **924**, 271 (2013).
[4]M. Moradi, V. Babin, C. Sagui, and C. Roland, Methods Mol. Biol. **924**, 313 (2013).
[5]T. Ichiye and M. Karplus, Proteins: Struct., Funct., Bioinf. **11**, 205 (1991).
[6]G. Huber, Am. Math. Monthly **89**, 301 (1982).
[7]R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[8]A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, Proc. Natl. Acad. Sci. U.S.A. **98**, 5578 (2001).
[9]T. Strunk, M. Wolf, M. Brieg, K. Klenin, A. Biewer, F. Tristram, M. Ernst, P. J. Kleine, N. Heilmann, I. Kondov, and W. Wenzel, J. Comput. Chem. **33**, 2602 (2012).
[10]A. Verma and W. Wenzel, Biophys. J. **96**, 3483 (2009).
[11]K. V. Klenin and W. Wenzel, J. Chem. Phys. **135**, 235105 (2011).
[12]E. A. Coutsias, C. Seok, M. P. Jacobson, and K. A. Dill, J. Comput. Chem. **25**, 510 (2004).
[13]Y. G. Mu, P. H. Nguyen, and G. Stock, Proteins **58**, 45 (2005).
[14]G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[15]P. L. Freddolino, S. Park, B. Roux, and K. Schulten, Biophys. J. **96**, 3772 (2009).
[16]S. Park, A. Y. Lau, and B. Roux, J. Chem. Phys. **129**, 134102 (2008).
[17]W. J. Becktel and J. A. Schellman, Biopolymers **26**, 1859 (1987).
[18]J. H. Chen, C. L. Brooks, and J. Khandogin, Curr. Opin. Struct. Biol. **18**, 140 (2008).