Accurate Calculation of Conformational Free Energy Differences in Explicit Water: The Confinement–Solvation Free Energy Approach

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MARCH 2015

Impact Factor: 3.3 · DOI: 10.1021/acs.jpcb.5b01632 · Source: PubMed

2 AUTHORS:

Jeremy Esque
University of Strasbourg
8 PUBLICATIONS **26** CITATIONS

SEE PROFILE

Marco Cecchini
University of Strasbourg
44 PUBLICATIONS **1,099** CITATIONS

SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

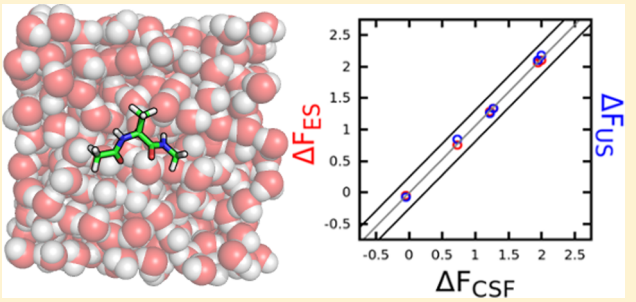
Accurate Calculation of Conformational Free Energy Differences in Explicit Water: The Confinement−Solvation Free Energy Approach

Jeremy Esque and Marco Cecchini*

Laboratoire d'Ingénierie des Fonctions Moléculaires (ISIS), UMR 7006 CNRS, Université de Strasbourg, F-67083 Strasbourg Cedex, France

Supporting Information

ABSTRACT: The calculation of the free energy of conformation is key to understanding the function of biomolecules and has attracted significant interest in recent years. Here, we present an improvement of the confinement method that was designed for use in the context of explicit solvent MD simulations. The development involves an additional step in which the solvation free energy of the harmonically restrained conformers is accurately determined by multistage free energy perturbation simulations. As a test-case application, the newly introduced confinement/solvation free energy (CSF) approach was used to compute differences in free energy between conformers of the alanine dipeptide in explicit water. The results are in excellent agreement with reference calculations based on both converged molecular dynamics and umbrella sampling. To illustrate the general applicability of the method, conformational equilibria of met-enkephalin (5 aa) and deca-alanine (10 aa) in solution were also analyzed. In both cases, smoothly converged free-energy results were obtained in agreement with equilibrium sampling or literature calculations. These results demonstrate that the CSF method may provide conformational free-energy differences of biomolecules with small statistical errors (below 0.5 kcal/mol) and at a moderate computational cost even with a full representation of the solvent.



INTRODUCTION

Biomolecular machines such as enzymes, motors, and switches perform a wide range of essential functions in the cell by cycling thorough a series of distinct conformational states, which are stabilized by chemical events.[1] In solution, these states are in equilibrium and the probability of finding the biomolecule in a given conformation is related to its free energy, which is proportional to the natural logarithm of the partition function of the associated conformational ensemble.[2] The binding of a ligand or its chemical transformation into one or more reaction products may significantly change the free energy of the protein−ligand complex, possibly shifting the equilibrium toward one conformer or the other.[3] A cute illustration of such a mechanism is provided by ligand-gated ion channels where the structural interconversion between the closed-channel and the open-channel forms is regulated by the binding of a neurotransmitter, which results into an ion flux through the postsynaptic membrane.[4] Providing means to determine the free energy of conformation accurately and from first-principles, i.e., using high-resolution structures and physics-based atomistic models, will offer a quantitative understanding of biomolecular function[5] and open up to ground-breaking applications from the design of allosteric modulators with tunable activity[6] to the elucidation of the energy-storage mechanism in motor proteins.[7]

The calculation of the free energy of conformation by computer simulations is a challenging task.[8] First, the functional conformational transitions in biomolecules typically occur in the millisecond to second time scales, which are not directly accessible by atomistic molecular dynamics. Second, molecular modeling at these size scales is bound to empirical force fields, whose inaccuracies may introduce systematic errors at best on the order of 1 kcal/mol.[9] Third, methods like thermodynamic integration[10] and the exponential formula,[11] which have been increasingly successful in alchemical free-energy simulations,[12] are not satisfactory for many conformational problems particularly those mediated by complex isomerization paths. Last, the use of an explicit treatment of the solvent, which was shown to significantly improve the quality of the simulation results,[13] is technically more involved and not yet mainstream. It follows that for an accurate determination of the free energy of conformation the development of efficient strategies able to work in the context of explicit solvent simulations is compelling. In addition, the constantly increasing power of computer resources[14] along with the recent improvements in the performance of the MD codes[15,16] motivate the effort in this direction.

The largest family of current free energy methods for conformational problems relies on umbrella sampling,[17] which enhances the efficiency of the calculation by biasing the

Received: February 17, 2015
Revised: March 24, 2015

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

interconversion between distinct conformational states through the use of restraining potentials. In combination with the weighted-histogram analysis method (WHAM),[18] this technique provides one- or two-dimensional potentials of mean force (PMF) over the chosen reaction coordinate(s), which can be used to quantify differences in conformational free energy.[19,20] These free energy calculations are statistically efficient, computationally inexpensive, and have been recently applied in the context of explicit solvent simulations of biomolecules.[21] Nonetheless, the difference in free energy is evaluated by driving the system over one or more barriers, such that both the efficiency and the accuracy of the calculations will strongly depend on the choice of the reaction path. Since finding a low-energy transformation with no *a priori* knowledge is challenging,[22] the choice of the reaction coordinates is often critical and results in one of the classic limitations of umbrella sampling. In addition, the resulting PMF may suffer from systematic errors due to insufficient sampling of the degrees of freedom orthogonal to the reaction coordinate(s),[23] which are hard to be detected.

For cases in which suitable reaction coordinates are not known, path-independent free energy methods offer a valuable alternative. Metadynamics[24] and deactivated morphing[25] are prominent examples of this family and have been successfully applied to conformational problems in explicit water.[26,27] Another approach is the hypothetical scanning method,[28] which has been applied to peptide chains immersed in a box of explicit water.[29] Within this family, the confinement method,[30,31] which is related to Einstein's early work on crystals[32] and later developments,[33,34] aims at the free energy of conformation by restraining the actual conformational ensembles to the harmonic state of reference, whose free energy is known analytically. Because the harmonic state corresponds to a fraction of the unbiased ensemble, the free energy of confinement can be computed accurately and efficiently with no *a priori* knowledge of the reaction path(s).[31] Recently, the confinement method has been successfully applied to medium-sized protein molecules such as the converter of myosin VI[35] and several chameleon sequences[36] with an implicit treatment of the solvent. Also, its most recent formulation,[37] which requires no matrix diagonalization on the calculation of the reference free energy, is particularly suited to deal with proteins with 10^4 atoms or more.

In this report, we present a new development of the confinement method, which has been designed for use in the context of explicit-solvent simulations. This variant, which is referred to as the confinement/solvation free energy (CSF) method, aims at the free energy of conformation by transforming the actual conformational states in solution into harmonic states in a vacuum through confinement simulations followed by hydration free-energy calculations. The performances of the newly introduced CSF approach are tested on the alanine dipeptide in a bath of explicit water and benchmarked against reference calculations based on both converged molecular dynamics and umbrella sampling. Applications to more complex peptides such as met-enkephalin (5 aa) and deca-alanine (10 aa) follow. The theory, the molecular systems, and the simulation setup are described first. The CSF results on the alanine dipeptide, met-enkephalin, and deca-alanine (10 aa) are then presented along with an analysis of the efficiency and the accuracy of the calculations. A possible implementation of the CSF method on the highly scalable molecular dynamics software NAMD is also described. A discussion on the significance of the results is given in the ending section.

MATERIALS AND METHODS

Theory. *The Confinement–Solvation Free Energy Method.* The confinement/solvation free energy (CSF) method is an original development of the confinemenent approach,[30,31] which was designed to work in the context of explicit-solvent simulations. The method aims at the free energy of conformation by transforming each conformer in solution into a harmonic state in a vacuum, whose absolute free energy is exact. The overall strategy is illustrated in Figure 1. The

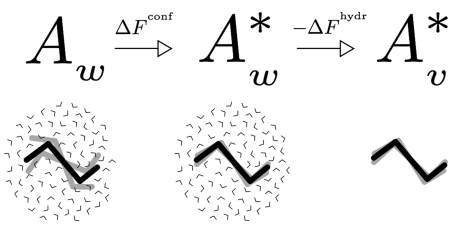


Figure 1. Confinement/solvation free energy (CSF) strategy.

actual conformational states (A_w) are first transformed into harmonic oscillators in a bath of explicit waters by means of restraining potentials with increasing strength. To complete the transformation to the reference state, the hydration free energy of the harmonically restrained conformers (A^*_w) is evaluated. As the absolute free energy of the harmonic state in a vacuum (A^*_v) is known analytically, the free energy of a given conformer (A_w) in solution is

$$F_{A,w} = F_{A^*,v} + \Delta F^{hydr}_{A^*} - \Delta F^{conf}_{AA^*} \quad (1)$$

where $\Delta F^{conf}_{AA^*}$ is the reversible work required to remove the nonharmonic contributions to the free energy of conformation and is obtained by confinement simulations;[30] $\Delta F^{hydr}_{A^*}$ is the hydration free energy of the harmonically restrained conformers, which is efficiently evaluated by free-energy perturbation (FEP) simulations;[38] and $F_{A^*,v}$ is the absolute free energy of $3N$ uncoupled harmonic oscillators in a vacuum, which is computed by normal-mode analysis.[39] Interestingly, eq 1 makes it clear that the free energy of conformation in solution can be obtained as a correction to the harmonic result in a vacuum via confinement simulations in explicit water to account for the anharmonicity and hydration free energy calculations to include the solvent contribution.

In this framework, the anharmonic contribution in eq 1, $\Delta F^{conf}_{AA^*}$, is evaluated by confinement simulations on the solute only, while leaving the solvent degrees of freedom unrestrained. By making use of the weighted-histogram confinement analysis (see below), the free energy of confinement is obtained as

$$\Delta F^{conf}_{AA^*} = F^{WHAM}_i \quad (2)$$

with F_i the value of the WHAM free-energy constant at the window corresponding to the strongest restraint strength, i.e., k_f in ref 30. The solvent contribution, $\Delta F^{hydr}_{A^*}$, is determined by driving alchemically the confined solute in a vacuum to a fully hydrated environment. By taking advantage of the WCA decomposition,[40] its hydration free energy is given by

$$\Delta F^{hydr}_{A^*} = \Delta F^{rep}_{A^*} + \Delta F^{disp}_{A^*} + \Delta F^{elec}_{A^*} \quad (3)$$

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

where the individual contributions are determined by multi-stage free-energy perturbation (FEP) simulations. Following Deng and Roux,[38] the repulsive contribution is evaluated using a nonlinear soft-core transformation, whereas linear coupling schemes are used both for the electrostatic and the dispersive contributions; see the Supporting Information for details.

Finally, the absolute free energy of the restrained conformers in a vacuum is obtained analytically using the classical formula for the canonical partition function of an ideal gas in the limit of the harmonic oscillator, rigid-rotor approximation.[39] Within these approximations, the molecular partition function has a closed form

$$z(V, T) = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} V \frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 IkT}{h^2}\right)^{3/2} \prod_i^{\kappa} \frac{kT}{h\nu_i} e^{D_e/kT} \quad (4)$$

with V being the volume, m the molecular mass, I the moment of inertia, σ the symmetry number, ν_i the vibrational frequencies, and D_e the energy of the ground electronic state, and the absolute free energy can be deconvoluted into translation, rotational, vibrational, and electronic contributions as

$$F_{A^*,v} = F^{tr}_{A^*,v} + F^{rot}_{A^*,v} + F^{vib}_{A^*,v} + F^{el}_{A^*,v} \quad (5)$$

which all have analytical expressions; see the Supporting Information.

Combining the confinement (eq 2), the hydration (eq 3), and the harmonic (eq 5) contributions to the free energy of conformation, the free energy change between conformers A and B in solution is obtained as

$$\Delta F_{CSF} = F_B - F_A = \Delta F_{A^*B^*} + \Delta\Delta F^{hydr}_{A^*B^*} - \Delta\Delta F^{conf}_{AB} \quad (6)$$

with $\Delta\Delta F^{conf}_{AB} = \Delta F^{conf}_{BB^*} - \Delta F^{conf}_{AA^*}$, $\Delta\Delta F^{hydr}_{A^*B^*} = \Delta F^{hydr}_{B^*} - \Delta F^{hydr}_{A^*}$, and $\Delta F_{A^*B^*} = F_{B^*,v} - F_{A^*,v}$.

In the special case of conformational equilibria, which is the focus of this paper, the analytical expression of $\Delta F_{A^*B^*}$ greatly simplifies and shows that the harmonic contribution only depends on the ground-state electronic energy and the vibrational frequencies of the harmonically restrained conformers. It follows that $\Delta F_{A^*B^*}$ can be evaluated by normal-mode analysis, i.e., using the molecular mechanics energy at the minimum and the vibrational frequencies from Hessian diagonalization in the presence of the strongest restraint, or by quasi-harmonic analysis of a room-temperature MD trajectory sampled at the strongest confinement. In the latter, the ground-state electronic energy is accessed from the ensemble-averaged potential energy $\langle V\rangle$ and the vibrational frequencies from the diagonalization of the mass-weighted covariance matrix;[41] see the Supporting Information for details.

Weighted-Histogram Confinement Analysis. In the original formulation of the confinement method,[30,31] the free energy of confinement was obtained from a large series of restrained simulations all resampling the same (harmonic) portion of the conformational basin. An alternative implementation, which is expected to improve the efficiency of the calculation, would be considering the confinement as a special case of umbrella sampling where the reference state is constant and the strength of the restraining potential is varied by window (Figure S4, Supporting Information). During the confinement, the work performed by the restraining potential is

$$w_i(\xi) = k_i(\xi - \xi_0)^2 \quad (7)$$

where $\xi - \xi_0$ measures the deviation from the reference structure and k_i the strength of the harmonic restraint in the ith window; note that k_i corresponds to λk_f in the original implementation of the confinement method.[30] In this special case of umbrella sampling, the reaction coordinate is the configurational distance from the reference, which is related to the root-mean-square deviation (RMSD) by

$$(\xi - \xi_0)^2 = N_{at}\mathrm{RMSD}^2 \quad (8)$$

with N_{at} being the number of atoms; see ref 31. Given an expression for the microscopic work of the biasing potential (eq 7) and a series of windows with restraining potentials of increasing strength (i.e., the confinement runs), the unbiased probability distribution over the reaction coordinate, $p^\circ(\xi)$, and the set of free energy constants, $\{F_i\}$, can be optimally estimated using WHAM.[18] As the free energy constants per window quantify the reversible work associated with the introduction of the biasing potential in each window, the entire set $\{F_i\}$ defines a potential of mean force along the restraint strength, k. Thus, the value of the free energy constant corresponding to the highest restraint strength corresponds to the free energy of confinement determined by thermodynamic integration.[30] The advantage of a weighted-histogram confinement analysis is that the free energy of confinement is obtained by optimally combining data collected from all windows and with no need for numerical integration.

Statistical Errors. In a typical CSF calculation (eq 6), only the confinement and the hydration free-energy contributions are subjected to statistical errors because the absolute free energy of the harmonically restrained conformers in a vacuum is exact. As such, the statistical error on the value of ΔF between conformers A and B can be estimated as

$$\delta\Delta F_{AB} = \sqrt{(\delta\Delta F^{conf}_{AB})^2 + (\delta\Delta F^{hydr}_{A^*B^*})^2} \quad (9)$$

which accounts for error propagation. Similarly, statistical errors on the confinement and the hydration free-energy contributions, which are evaluated as differences between pairs of independent free-energy measurements per conformer, are given by

$$\delta\Delta\Delta F^{conf}_{AB} = \sqrt{(\delta\Delta F^{conf}_{BB^*})^2 + (\delta\Delta F^{conf}_{AA^*})^2} \quad (10)$$

and

$$\delta\Delta\Delta F^{hydr}_{AB} = \sqrt{(\delta\Delta F^{hydr}_{B^*})^2 + (\delta\Delta F^{hydr}_{A^*})^2} \quad (11)$$

where $\delta\Delta F^{conf}$ and $\delta\Delta F^{hydr}$ are the standard errors on the confinement and the hydration free energies of conformers A and B.

When the harmonic contribution is evaluated by QHA, the ground-state electronic energy and the vibrational frequencies are determined from the ensemble-averaged potential energy and the eigenvalues of the covariance matrix obtained from a room-temperature simulation trajectory in the presence of the strongest harmonic restraint; see the Supporting Information. Although both quantities are in principle subjected to statistical errors, because of the strong confinement, the vibrational frequencies converge rapidly and can be considered as exact. If so, statistical errors in the QHA calculation may arise only from

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

the ensemble-averaged potential energy and can be estimated from the standard error of the room-temperature distribution as

$$\delta F^{\mathrm{QHA}}_{\mathrm{A}^*,\mathrm{v}} = \frac{\sigma(V)}{\sqrt{M}} \tag{12}$$

with V being the potential energy and M the number of snapshots considered for the ensemble average. In the QHA version of CSF, the error in eq 12 must be included in eq 9 upon accounting for error propagation as done for the confinement and the hydration free energy contributions.

Systems and Setup. The molecular system chosen for the validation of the CSF method is the *N*-acetyl-*N*′-methylamide derivative of alanine, commonly referred to as alanine dipeptide, which has become a standard model for the theoretical investigation of conformational equilibria in biomolecules.[24,31,42−45] Then, to demonstrate the applicability of the new development in the broader context of biomolecules, two larger folding peptides, i.e., met-enkephalin (5 aa) and deca-alanine (10 aa), were considered.

Alanine Dipeptide. The chemical structure of the dipeptide is shown in Figure 2A. The molecule involves 22 atoms including hydrogens and is highly flexible due to the "soft" dihedrals of the backbone. Indeed, at room temperature and in solution, the dipeptide is in rapid equilibrium between multiple conformers, which can be entirely described in terms of the dihedral angles ϕ and ψ. Four distinct free-energy basins can be identified in explicit-water simulation at 300 K; see Figure 2B. They correspond to the extended conformation of the backbone (c7eq), the right-handed (α_R) and left-handed (α_L) α-helical conformers, and the γ_D-turn configuration of the backbone (c7ax) as per Ramachandran's terminology;[46] geometrical definitions of these basins based on the values of ϕ and ψ are given in Table S1 (Supporting Information). Clearly, the backbone dihedrals (ϕ, ψ) identify natural "reaction coordinates" to analyze the conformational transitions of the dipeptide.

Met-Enkephalin. Met-enkephalin is a five-residue peptide (YGGFM) acetylated at the N-terminus and amidated at the C-terminus, which is known to inhibit the release of neurotransmitter upon activation of different opioid receptors.[47] The ability of binding to several targets shows that this peptide may adopt distinct conformations in solution. Indeed, despite its reduced size, i.e., 81 atoms including hydrogens, met-enkephalin has become a prototypical example of a difficult case for sampling.[48,49] Among all possible conformers, two of them were considered for a free energy analysis by CSF. They correspond to the helical 3_{10} and the γ-turn configurations of the backbone (Figure 3A), which are referred to as conformers I and IV following the terminology used in ref 49. Initial coordinates for the simulations were obtained from the first model of the NMR structure 1PLW[50] available in the PDB.

Deca-Alanine. Deca-alanine is a 10-residue peptide made of consecutive alanines and blocking groups (112 atoms including hydrogens). Despite the extensive literature on this peptide, which has been used as a model system for (peptide) folding,[51−56] free energy analyses with an explicit treatment of the solvent are rare.[25,57] Here, we consider the α-helical to β-hairpin transition (Figure 3B) and have evaluated the free energy change in solution using the CSF method. Initial coordinates for the α state were obtained from ref 58. Those for the β state were obtained from the NMR structure of a *de novo* designed β-hairpin peptide (1J4M) upon mutation of its amino-acid residues from 3 to 12 into alanines. For a similar conformational transition of deca-alanine in explicit water, a value of 10 ± 1 kcal/mol in favor of the α-helical structure was reported previously.[25]

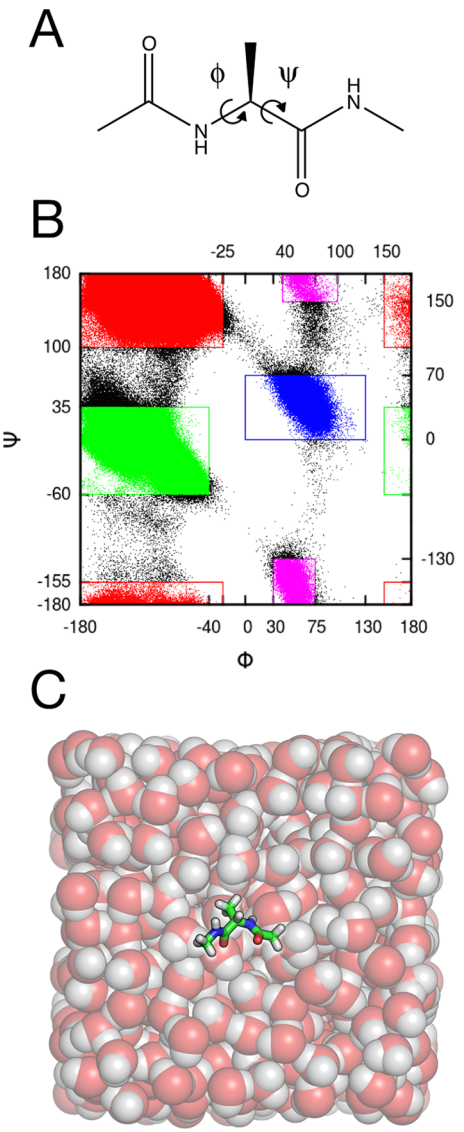


Figure 2. Alanine dipeptide. (A) The chemical structure of the dipeptide. The soft dihedral angles of the backbone (ϕ, ψ) are indicated by arrows. (B) The conformational basins of the dipeptide in solution at 300 K shown as nonoverlapping regions of the Ramachandran plot: c7eq (red), α_R (green), α_L (blue), and c7ax (magenta). (C) The cubic box to simulate the dipeptide in a bath of explicit water. The dipeptide is represented in sticks, the water molecules by van der Waals spheres.

Simulation Setup. The atomistic model of the alanine dipeptide was built using the CHARMM22 force field[59] with CMAP corrections for the backbone.[60] The dipeptide was embedded in a cubic box of 566 water molecules pre-equilibrated in the NPT ensemble at 300 K. The solvent was simulated using a modified TIP3P water model[61] and imposing periodic boundary conditions. The short-range nonbonded interactions were truncated using a distance cutoff of 10 Å and smoothly switched to zero from 9 to 10 Å. Long-range electrostatic interactions were accounted for by the particle mesh Ewald (PME) method using a grid spacing of 0.9 Å and a sixth-order B-spline charge interpolation scheme.[62] Covalent

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

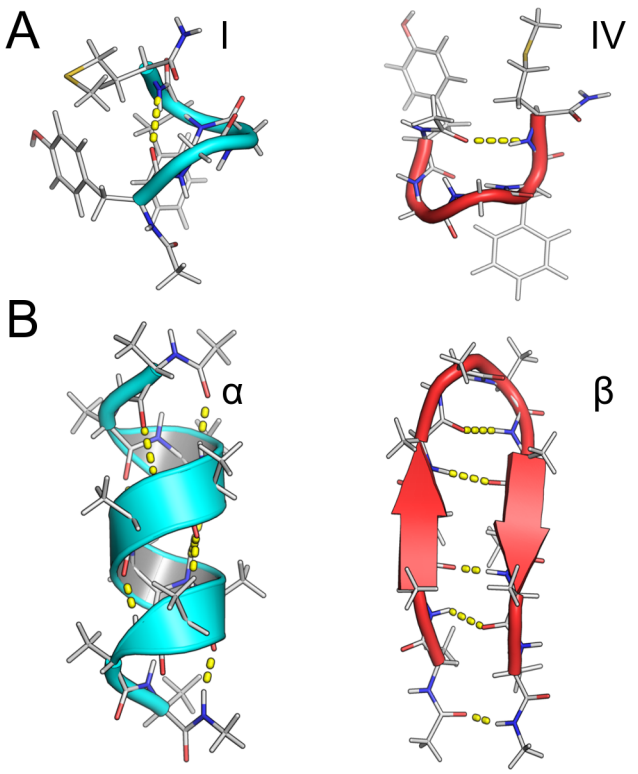


Figure 3. Folding peptides. (A) Reference structures for met-enkephalin with the 3_{10} helical motif (I) and the γ-turn (IV) conformers shown on the left- and right-hand sides, respectively. (B) Reference structures for deca-alanine with the α-helical (α) and β-hairpin (β) conformers shown on the left- and right-hand sides, respectively. Yellow dashed lines highlight the H-bonding interactions.

bonds involving the hydrogen atoms were left unconstrained throughout the simulations; i.e., no SHAKE[63] or equivalent constraint was applied. All calculations were performed using CHARMM.[64]

Starting with a fully solvated configuration of the dipeptide in c7eq, the system was energy minimized by 200 steps of steepest descent (SD) followed by 500 steps of the adopted basis Newton−Raphson (ABNR) optimization. The resulting configuration was heated up from 100 to 300 K using a linear temperature gradient while imposing an external pressure of 1 atm. Finally, the system was equilibrated in the isothermal−isobaric (NPT) ensemble for 500 ps at a constant temperature of 300 K using a time step of 1 fs. During the simulation, pressure control was achieved using the extended system of Andersen[65] and setting the mass of the piston to 500 amu. Temperature was controlled by coupling to a Nosé−Hoover thermostat[66,67] with the mass of the piston set to 1000 kcal/mol/ps^2. The equilibration in the NPT ensemble produced a stable trajectory over time, which was used to extract a smaller box of 25.55 Å per side around the protein; see Figure 2C. The resulting configuration was shortly equilibrated (15 ps) in the canonical ensemble (NVT) and used as a starting structure for all subsequent simulations.

Atomistic models for met-enkephalin and deca-alanine were built using the CHARMM22 force field[59] and CMAP corrections.[60] Both peptides, which are neutral in charge, were embedded in a cubic box of pre-equilibrated water with periodic boundaries including 779 and 3436 water molecules for met-enkephalin and deca-alanine, respectively. Apart from a PME grid spacing of 1 Å, the use of a Berendsen barostat[68] and a Langevin thermostat[69] to sample the NPT ensemble, the simulation setup was the same as the one used for the alanine dipeptide. All calculations were performed with the program NAMD 2.9.[15] For met-enkephalin, an initial 5000 steps of energy minimization of the fully solvated system was performed starting with the coordinates of the NMR structure (1PLW). The resulting structure was heated up from 2 to 300 K using a linear temperature gradient and equilibrated at a temperature of 300 K for 200 ps, while imposing an external pressure of 1 atm. A cubic box of 28.8 Å per side around the peptide was extracted from the equilibrated trajectory in the NPT ensemble and used as the starting structure for the subsequent NVT simulations. A similar procedure was followed to produce an equilibrated water box of 48.0 Å per side for both α and β conformers of deca-alanine.

Reference Calculations. *Equilibrium Sampling.* The great advantage of the alanine dipeptide is that given the small molecular size (22 atoms) and the low barriers separating the conformational states in solution (<6 kcal/mol) reference values for ΔF can be accessed by converged MD simulations or equilibrium sampling (ES). For this purpose, μs-long trajectories of the dipeptide in explicit water were collected in the NVT ensemble. After clustering molecular configurations based on the dihedral angles of the backbone (Table S1, Supporting Information), the populations of the four conformers at equilibrium were used to estimate differences in free energy as

$$\Delta F_{ES} = F_B - F_A = -kT \log(N_B/N_A) \quad (13)$$

where N_A and N_B are the number of snapshots of conformers A and B sampled by the simulations, T is the temperature in Kelvin, and k is the Boltzmann constant.

Umbrella Sampling. The availability of natural reaction coordinates to describe the conformational transitions of the dipeptide, i.e., the ϕ and ψ backbone dihedrals, makes it possible to obtain accurate estimates of ΔF by umbrella sampling (US). By dividing the dihedral space of the dipeptide into windows and carrying out a series of room-temperature simulations in the presence of dihedral restraints, US in conjunction with WHAM[18] provides an efficient strategy for accessing a two-dimensional PMF in the dihedral space. Starting with the unbiased probability distribution per bin and using the same definitions for the basins, values of ΔF for all pairs of conformers were obtained as

$$\Delta F_{US} = F_B - F_A = -kT \log \frac{\sum_{j\in B} p_j}{\sum_{i\in A} p_i} \quad (14)$$

where p_i and p_j are the unbiased probabilities of the dihedral bins i and j, which belong to the conformational basins A and B, respectively; see the Supporting Information for a derivation of eq 14.

RESULTS AND DISCUSSION

Equilibrium Sampling. Equilibrium simulations of the alanine dipeptide in a bath of explicit water were carried out in the NVT ensemble using the setup described above. For this purpose, three independent MD runs of 1 μs at 300 K were produced starting with an equilibrated configuration of the system and random initial velocities. The trajectories were analyzed by computing the time series of the dihedral angles (ϕ, ψ), which provide estimates of the populations of the

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 1. Conformational ΔF between States of the Alanine Dipeptide Computed from Equilibrium Sampling (ES), Umbrella Sampling (US), and the Confinement/Solvation Free Energy Method (CSF)[a]

conformers	ΔF_{ES}	ΔF_{US}	ΔF_{CSF}	ΔF_{CSF}^{NAMD}
c7eq/α_R	−0.05 ± 0.02	−0.08 ± 0.02	−0.05 ± 0.04	−0.07 ± 0.08
c7eq/α_L	1.29 ± 0.04	1.26 ± 0.04	1.22 ± 0.05	1.32 ± 0.08
c7eq/c7ax	2.06 ± 0.05	2.10 ± 0.04	1.96 ± 0.05	2.10 ± 0.08

[a]All values are given in kcal/mol.

various conformers at equilibrium. Differences in conformational free energy were then obtained using eq 13. The equilibrium sampling (ES) results in Table 1 show that the c7eq and α_R conformers correspond to the deepest free-energy minima, which are essentially equivalent, whereas α_L and c7ax are relative free-energy minima higher by 1 or 2 kcal/mol, respectively.

Statistical Errors. The uncertainties on the ΔF values computed by ES were estimated by bootstrapping[70] after separating the trajectories into "round trips"; here with round trip we refer to the fraction of contiguous snapshots along the trajectory in which the system starts from basin A, visits basin B, and returns to A (Figure S5, Supporting Information). Upon merging the MD runs, a total of 3350 round trips between c7eq and α_R, 480 round trips between c7eq and α_L, and 282 round trips between c7eq and c7ax were collected. Provided that the geometric definitions of the free energy basins preserve the transition barrier(s) between conformers,[71] conformational round trips decompose the trajectory into a series of independent events and yield a collection of statistically independent measurements of ΔF_{AB}. Using the round-trip data sets, unbiased estimates of $\delta\Delta F$ were obtained from the standard error of the bootstrapping distribution obtained by drawing randomly and with replacement N independent measurements of ΔF, with N being the total number of round trips; see the Supporting Information. The results are given in Table 1. For the alanine dipeptide in explicit water, ES provides extremely precise conformational free energy estimates with statistical errors well below the "chemical accuracy",[72] i.e., $\delta\Delta F < 0.1$ kcal/mol.

By varying the number of round trips per resample (N), the bootstrapping results show that the magnitude of $\delta\Delta F$ is linear with $1/(N)^{1/2}$ independently of the pair of conformers and with a slope close to 1; see Figure 4. This result indicates that the number of round trips sampled during the trajectory (both forward and backward transitions) solely determines the magnitude of the uncertainty on ΔF, which implies that the convergence of ΔF in ES is rate-determined by the height of the barrier separating the conformational basins. Moreover, considering that the average slope is 0.9, the correlation in Figure 4 suggests that three round trips would be sufficient to obtain an estimate of ΔF with a statistical error of 0.5 kcal/mol. On the other hand, approximately 100 round trips would be needed to obtain results with "chemical accuracy", i.e., $\delta\Delta F <$ 0.1 kcal/mol. Interestingly, because the slope in Figure 4 is essentially independent of the value of ΔF, i.e., it is the same for c7eq/α_R, c7eq/α_L, and c7eq/c7ax, these conclusions are expected to be valid for systems of greater complexity.

Umbrella Sampling. Umbrella sampling (US) of the dihedral angles of the backbone ϕ and ψ was performed to obtain orthogonal estimates of the ΔF between conformers of the alanine dipeptide in explicit water. For this purpose, the configurational space of each dihedral was divided into 12 windows (i.e., a 30° resolution) and 144 MD simulations, each

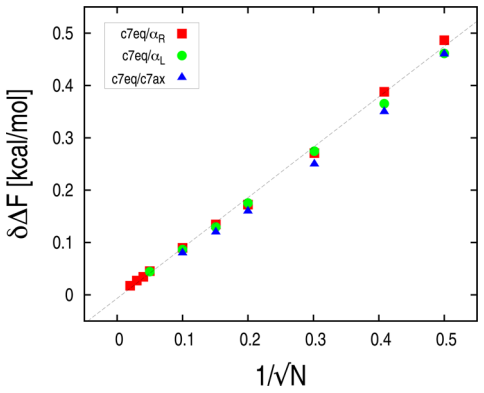


Figure 4. Evaluation of the statistical error on ΔF from molecular dynamics at equilibrium. Errors were computed by bootstrapping conformational round trips as described in the main text. The data show that $\delta\Delta F$ is linear with $1/(N)^{1/2}$, with N being the number of round trips sampled during the simulation trajectory. The slopes of the linear fits for c7eq/αR, c7eq/αL, and c7eq/c7ax are 0.94, 0.90, and 0.87, respectively.

of 4 ns in length, were carried out at 300 K in the presence of dihedral restraints. The strength of the restraint was set to 10 kcal/mol/rad^2 for all values of ψ and $-180° \le \phi < -30°$. Stronger restraints, i.e., k of 20 kcal/mol/rad^2, were used for $-30° \le \phi \le 180°$ to improve the convergence of sampling. After a short equilibration of 100 ps, the time series of the backbone dihedrals were collected over 4 ns per window (i.e., 4000 snapshots) and processed by WHAM[18] using the software available from Grossfield's Web site (http://membrane.urmc.rochester.edu/content/wham). Then, the unbiased probability distributions of ϕ and ψ computed by WHAM were used to construct the two-dimensional PMF shown in Figure 5. The PMF was determined with a resolution of 4° on both dihedral coordinates taking into account periodicity and using a convergence criterion of 1×10^{-5} kcal/mol on the value of the free energy constants between any two consecutive iterations. On the basis of the definitions in Table S1 (Supporting Information), conformational free-energy differences between the various conformers were obtained using eq 14. The free energy results (US) for the alanine dipeptide in explicit water are given in Table 1. For a total sampling time of 576 ns, the US results are in quantitative agreement with the ES analysis.

Statistical Errors. The statistical uncertainty on the US results was estimated by bootstrapping molecular snapshots from the trajectories sampled by the various umbrella windows. As discussed in ref 74, resampling trajectories is a useful strategy to produce a large number (N_{boot} = 1000) of statistically independent umbrella sampling at no additional cost. These resamples were processed by WHAM (see above) and used to obtain an equivalent number of independent estimates of ΔF for each conformational pair. Unbiased estimates of $\delta\Delta F$ were then obtained from the standard error

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

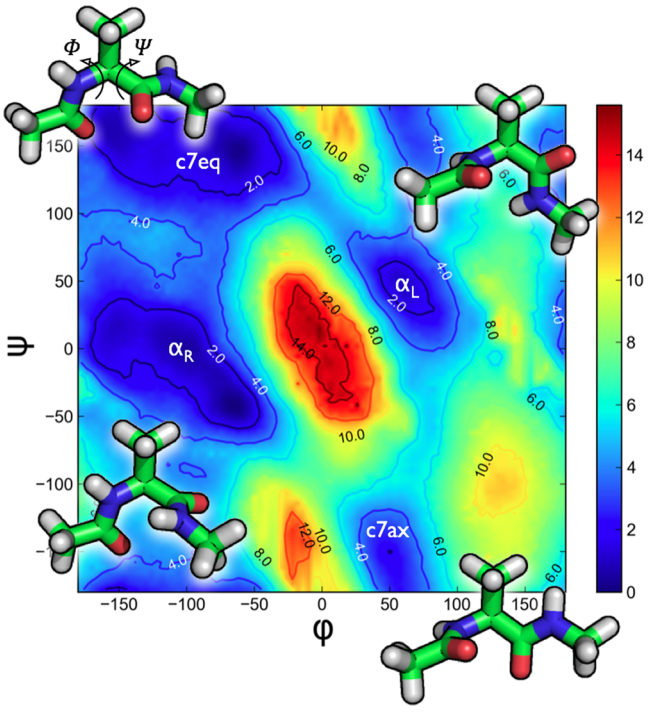


Figure 5. Conformational free energy surface of the alanine dipeptide in explicit water. The two-dimensional potential of mean force (PMF) was obtained from umbrella sampling (US) of the backbone dihedrals ϕ and ψ. The PMF shows the existence of four conformational basins (c7eq, α_R, α_L, and c7ax) whose representative structures are shown in sticks. Contour lines which represent ISO-energetic levels are given in kcal/mol. Molecular representations were drawn with the help of PyMol.[73]

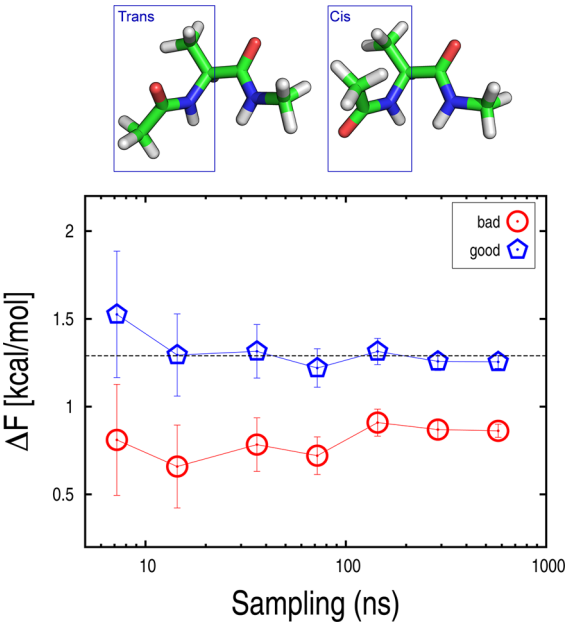


Figure 6. Systematic errors in umbrella sampling. Convergence of ΔF between the c7eq and α_L conformers of the alanine dipeptide for the "bad" (red) and "good" (blue) PMF; see the main text. Statistical errors on the value of ΔF, which were estimated by bootstrapping, are shown as error bars. The black dashed line corresponds to the reference ΔF from ES, i.e., 1.29 kcal/mol. On top, a pictorial representation of the *trans* to *cis* isomerization is given. In the *trans* configuration, the nitrogen and oxygen atoms of the same peptide group point in opposite directions, unlike in the *cis* configuration.

of the bootstrapping distribution; see Figure S6 (Supporting Information). The results are given in Table 1.

Systematic Errors. As illustrated by Zhu and Hummer,[23] free-energy calculations based on umbrella sampling and WHAM may suffer from systematic errors when sampling on the degrees of freedom orthogonal to the reaction coordinate(s) is inadequate. The occurrence of such systematic errors is exemplified here by comparing two umbrella sampling calculations of the dipeptide in explicit water. In the first calculation, which is referred to as "bad" PMF, umbrella sampling along ϕ and ψ was performed by applying the window-dependent biasing potential directly to a pre-equilibrated structure of the c7eq conformer in a bath of explicit water at 300 K. After discarding the first 100 ps of simulations, which were required to drive the system to the appropriate region of the Ramachandran plot, the time series of the dihedral angles were collected and used to produce the two-dimensional PMF. The difference in free energy between c7eq and α_L determined from the resulting PMF is given in Figure 6 as a function of the total simulation time (sampling). Surprisingly, the value of ΔF converged to 0.8 kcal/mol, which is off by half kcal/mol relative to the ES estimate for the same pairs of conformers; see Table 1. A careful analysis of the simulation trajectories in the various windows revealed that in 15 over 144 windows the abrupt introduction of the biasing potential produced a *trans* to *cis* isomerization of one of the two Ω dihedrals, i.e., the dihedral angle of the backbone corresponding to a rotation about the peptide bond (Figure 6, top). Since these transitions are forbidden at room temperature, the *trans* to *cis* isomerization corresponds to an irreversible transition orthogonal to the ϕ and ψ reaction coordinates. Introducing additional restraints on the Ω angles during the equilibration of the various windows prevented the *trans* to *cis* isomerization and provided a correct PMF, which is referred to as "good" in Figure 6. Correcting for this artifact results in a free energy estimate that is in quantitative agreement with the ES results. Given the smooth convergence of the standard error on ΔF with increasing sampling (error bars in Figure 6), the occurrence of systematic errors in these calculations is hardly detectable. More generally, any slow-relaxing degree of freedom orthogonal to the reaction coordinate(s) used in US may introduce systematic errors, which will contaminate the free-energy calculation results.

Confinement–Solvation Free Energy Method. The results obtained from ES and US on the alanine dipeptide were used to benchmark the performance of the CSF method. As described above, CSF aims at the free energy of conformation by transforming the molecular conformers in solution into harmonic states in a vacuum. To this aim, the actual conformers of the dipeptide were first transformed into harmonic oscillators in a bath of explicit water by confinement simulations.[30,31] The transition to in a vacuum was then accomplished through an alchemical transformation relying on multistage FEP simulations.[38] For the confinement, 15 restrained simulations in explicit water were carried out with harmonic force constants from 0.005 to 82 kcal/mol/Å^2 with positional restraints applied to the peptide atoms only; the actual values of k were 0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.2, 20.5, 40.9, and 82 kcal/mol/Å^2. Following the prescriptions given in ref 31, the confinement was done using a BESTFIT harmonic restraint on a reference configuration of the dipeptide, which was energy-minimized in

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

the presence of the solvent. In detail, reference configurations for the various confinement calculations were produced by explicit-solvent MD relaxations with dihedral restraints to guide the dipeptide to the basin of interest, followed by a short equilibration in the absence of restraints and energy minimization by 10 000 steps of steepest descent with a tolerance gradient of 1.0×10^{-9} kcal/mol/Å; the values of ϕ and ψ of the reference configurations are given in Table S2 (Supporting Information). For the confinement, each restrained simulation was 40 ns long (i.e., 40 000 snapshots), yielding a total sampling of 600 ns per conformer. By using the same dihedral definitions for the basins (Table S1, Supporting Information) to select representative configurations from the various runs and processing those by a weighted-histogram confinement analysis (see the Materials and Methods section), confinement free energies ranging from 20.43 to 21.24 kcal/mol were obtained for the four conformers (Table S3, Supporting Information). Corresponding statistical errors were computed by bootstrapping, as described in the Supporting Information. As shown in Table S4 (Supporting Information), these results are equivalent to those obtained with thermodynamic integration but the corresponding errors are 3−6 times smaller. Thus, the weighted-histogram analysis appears to improve the efficiency of the calculations by a factor of 3. Data collected from the restrained runs with k lower than 0.005 kcal/mol/Å^2 had no effect on the free energy results (Figure S8, Supporting Information) and were disregarded.

The hydration free energy of the harmonically restrained conformers was determined by multistage FEP simulations following the WCA scheme in Figure S1 (Supporting Information). All simulations were performed using the PERT module in CHARMM.[64] The contribution from the core repulsion, ΔF^{rep}, was determined by a multistage nonlinear soft-core transformation with the staging parameter s set to 0.0, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and 1.0. Because of the nonlinear coupling, each stage of the calculation was solved individually by evaluating the reversible work of turning on the core−core repulsive interactions stepwise through a linear coupling via the ζ parameter. For this purpose, two simulations of 500 ps in length with $\zeta = 0$ (initial state; $s = s_i$) and $\zeta = 1$ (final state; $s = s_{i+1}$) were performed for each value of s_i. By contrast, the free-energy contributions from both the dispersive, ΔF^{disp}, and electrostatic interactions, ΔF^{elec}, were computed using linear coupling schemes with λ set to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Similarly to the weighted-histogram confinement analysis, the potential of mean force over the corresponding coupling parameters (ξ and λ) was obtained from the set of free energy constants computed by WHAM. To this aim, samples collected from 500 ps trajectories (500 000 snapshots) in all windows were processed by WHAM using a convergence criterion of 0.001 kcal/mol on the value of free energy. Overall, estimates of the repulsive, dispersive, and electrostatic contributions to the hydration free energy (eq 3) were obtained from a series of 792 (396 ns), 242 (121 ns), and 242 (121 ns) λ-dynamics simulations. By combining these contributions, hydration free energies ranging from −15.88 to −17.39 kcal/mol were obtained. Corresponding statistical errors were estimated by block averaging over a series of 10 independent calculations.

Finally, the absolute free energy of the harmonic states was evaluated by normal-mode analysis of the restrained conformers in a vacuum with k = 82 kcal/mol/Å^2. As shown in Table S3 (Supporting Information), the harmonic free energies are large and vary from 58.97 and 61.00 kcal/mol. Although the absolute value of the harmonic free energy is not relevant without a quantum treatment of the high-frequency vibrations,[75] the difference between pairs of harmonic states is meaningful and in the presence of strong confinement mainly reflects the difference in potential energy between the two conformers at their minimum (Table S3, Supporting Information).

By combining contributions from the confinement, the hydration, and the harmonic free energy per conformer, the conformational ΔF in solution were estimated for all pairs of conformers using eq 6. The results for the alanine dipeptide are given in Table 1. The calculated ΔF are in quantitative agreement with those obtained by ES and US. The striking correlation with both ES and US results with a slope of 1.06 and a spread less than 0.1 kcal/mol (Figure 7) demonstrates

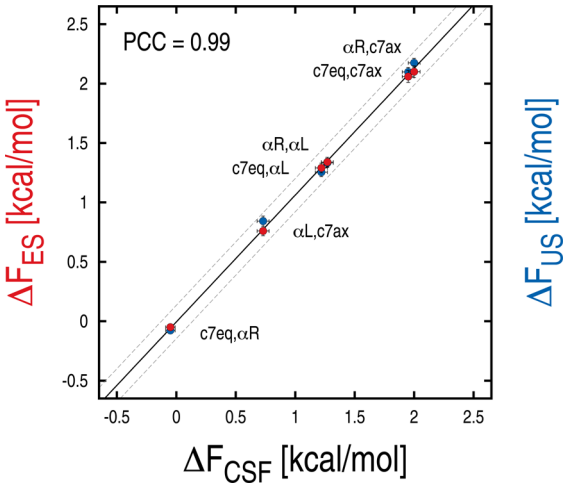


Figure 7. Correlation of the CSF results with those obtained from equilibrium sampling (ES) and umbrella sampling (US). Horizontal and vertical bars correspond to statistical errors on ΔF from CSF and ES/US, respectively. The strong correlation is demonstrated by a Pearson correlation coefficient (PCC) of 0.99. The linear fit of the free-energy results (solid line) provides a model with a slope of 1.06 and a Y-intercept of −0.006 kcal/mol. The dashed lines show that all data points are included in a range of ±0.1 kcal/mol from the linear fit.

the accuracy of the CSF calculations. Also, the CSF results show smooth convergence with increasing k (Figure S2, Supporting Information), which is a practical but important requirement to achieve control over the calculations. Finally, the decomposition of ΔF reported in Table 2 indicates that the solvent contribution to the free energy of conformation is significant and needs to be accounted for correctly. For instance, it appears that the thermodynamic stability of α_{L} versus α_{R} is significantly increased by its more favorable interaction with the solvent, despite being sensibly more strained and less anharmonic. Also, the ΔF between α_{R} and c7ax, which appears to be well captured by the harmonic analysis in a vacuum, is so because of the exact cancelation of the solvent and anharmonic contributions. The quantification of both the anharmonicity and the solvent contributions by CSF thus provides insights on the microscopic origin of conformational stability in solution.

Efficiency. The efficiency of the CSF method was evaluated by monitoring the variance of the calculated ΔF as a function of

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 2. Decomposition of the Conformational ΔF Predicted by CSF[a]

conformers A/B	ΔF_{AB}	$-\Delta\Delta F_{AB}^{conf}$	$\Delta\Delta F_{A^*B^*}^{hydr}$	$\Delta F_{A^*B^*}$
c7eq/α_R	−0.05 (±0.04)	−0.28 (±0.02)	0.47 (±0.04)	−0.24
c7eq/α_L	1.22 (±0.05)	0.53 (±0.02)	−1.04 (±0.05)	1.73
c7eq/c7ax	1.96 (±0.05)	−0.13 (±0.02)	0.30 (±0.04)	1.79
α_R/α_L	1.27 (±0.05)	0.81 (±0.02)	−1.51 (±0.05)	1.97
α_R/c7ax	2.01 (±0.05)	0.15 (±0.02)	−0.17 (±0.04)	2.03
α_L/c7ax	0.74 (±0.05)	−0.66 (±0.02)	1.34 (±0.05)	0.06
I/IV	−0.16 (±0.13)	4.21 (±0.05)	−0.49 (±0.11)	−3.88 (±0.04)
α/β	11.44 (±0.31)	−21.82 (±0.09)	−5.62 (±0.29)	38.88 (±0.06)

[a]The CSF free energy change (ΔF_{AB}) is the sum of the confinement ($-\Delta\Delta F_{AB}^{conf}$), the hydration ($\Delta F_{A^*B^*}^{hydr}$), and the harmonic ($\Delta F_{A^*B^*}$) contributions; see eq 6. Statistical errors on the individual components are given in brackets. Results for met-enkephalin (conformers I and IV) and deca-alanine (conformers α and β) were obtained with the NAMD implementation of CSF, which accesses the harmonic contribution by quasi-harmonic analysis. All values are given in kcal/mol.

sampling. For this purpose, the total simulation time required to achieve a statistical precision of 0.1 kcal/mol, which yields predictions with "chemical accuracy", was used as a quantitative measure in the comparison with reference calculations by ES and US. As the hydration free-energy contribution to the conformational ΔF accounts for 10% of the total sampling for the alanine dipeptide, the convergence of the CSF calculations was assessed from the analysis of the confinement contribution only. The convergence rates of ΔF for c7eq/α_R, c7eq/α_L, and c7eq/c7ax are compared in Figure 8. The first and last pairs correspond to transition paths characterized by the lowest (4.2 kcal/mol, c7eq/α_R) and highest (5.2 kcal/mol, c7eq/c7ax) free-energy barriers from c7eq, respectively. The transition to c7ax is rate limited by the interconversion with α_L, which thus represents an on-pathway intermediate; see the PMF in Figure 5. The results show that the US calculations are most efficient with convergence rates ranging from 40 to 100 ns for the three conformational pairs. The performance of ES is in the same range for c7eq/α_R (100 ns), but the convergence is significantly slower for both c7eq/α_L and c7eq/c7ax (500 ns). The CSF method is faster converging for c7eq/α_L and c7eq/c7ax with performances that are intermediate between US and ES, although it requires twice as much to yield an accurate value of ΔF for c7eq/α_R. Taken together, these results indicate that the efficiency of equilibrium sampling (ES) is very sensitive to the height of the interconversion barrier, so that an increase of ∼1 kcal/mol in the barrier results in a 5-fold increase in the convergence time of ΔF. In sharp contrast, the performances of the CSF approach, which are by definition barrier(s) independent, appear to be rate-limited by the complexity of the basins, i.e., the number of sub-basins separated by free-energy barriers larger than kT. Finally, the comparison with the US results shows that when idealized reaction coordinates are available, e.g., the backbone dihedrals for the alanine dipeptide, this strategy is 2−5 times more efficient than the CSF approach and should be preferred. Nonetheless, the CSF strategy has the considerable advantage of being path independent[30] and unlike US provides estimates of ΔF with no need for projections over order parameters, which are likely to introduce systematic errors, as shown by Figure 6. As such, for complex molecular systems whose reaction coordinates are unknown *a priori*, we expect CSF to be most robust to systematic errors. Finally, we note that the efficiency comparison in Figure 8 may look significantly different for more complex solutes and is likely to favor even more the CSF strategy, as we shall see for deca-alanine.

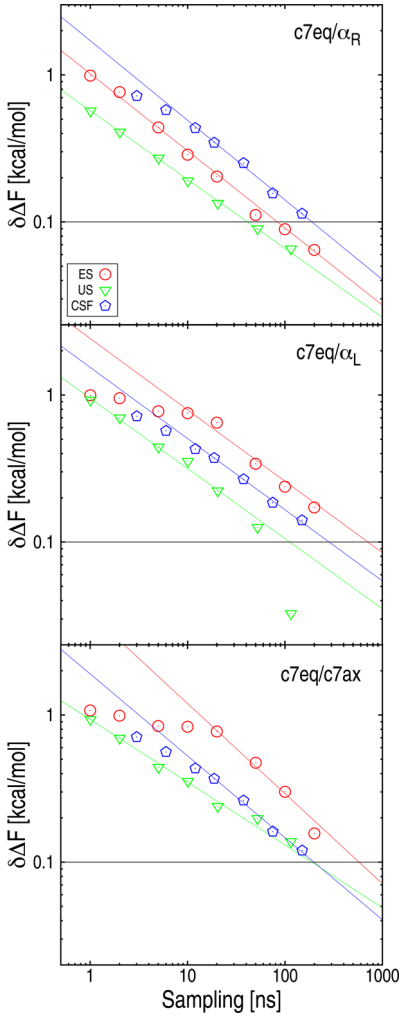


Figure 8. Efficiency of the CSF method (blue) relative to equilibrium sampling (red) and umbrella sampling (green). The statistical uncertainty on the difference in conformational free energy ($\delta\Delta F$) is given for the three methods as a function of sampling. Errors were estimated by block averaging as described in the Supporting Information. For CSF, only the convergence of the confinement contribution was considered. These results are to be complemented with the data in Figure S3 (Supporting Information), which show fine convergence of the three methods to the correct values of ΔF; i.e., no systematic error was present.

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Porting CSF to NAMD. To extend the scope of the new development to more complex solutes, it was crucial to have CSF implemented on highly scalable molecular dynamics software packages. In the following, a straightforward implementation of CSF on NAMD 2.9[15] is presented along with benchmark calculations on the alanine dipeptide.

Confinement simulations were performed using the *TMD* module by setting both the initial and target RMSD from the reference to 0 Å. The force constant of the restraining potential per window must be updated considering that restraint strengths in CHARMM correspond to $2N_{at}k$ in NAMD, with N_{at} being the number of atoms and k the spring constant in kcal/mol/Å^2. In addition, to ensure stability of the simulations, particularly in the high-force constant range, the Langevin damping constant was set to 10.0 ps^{-1} and the integration time step reduced to 0.5 fs. The hydration free energy of the harmonically restrained conformers was determined by FEP simulations within the *alchemical* module using the decoupling strategy described in ref 76. The main differences compared to CHARMM are: (i) both polar (electrostatic) and nonpolar (dispersive and repulsive) solute−solvent interactions were controlled by a single coupling parameter λ and scaled according to the coupling scheme shown in Figure S10 (Supporting Information); (ii) the solute−solute interactions were not annihilated during the alchemical transformation, which avoids running additional simulations in a vacuum; (iii) the nonpolar term was not split into repulsive and dispersive contributions. To maintain the solute near the center of the simulation box, an additional harmonic restraint with a force constant of 1 kcal/mol/Å^2 was applied to the center of mass of the solute. The hydration free energy and the corresponding statistical error were computed using the Bennett acceptance ratio (BAR) method as implemented in the ParseFep program.[77] Finally, because Hessian diagonalization is not implemented in NAMD, the absolute free energy of the harmonic states was accessed by quasi-harmonic analysis (QHA)[41] of NAMD trajectories collected in the presence of the strongest restraint in a vacuum, which yields identical results to NMA in the harmonic limit.[31] In the current implementation, molecular conformers in a vacuum were restrained using the COLVAR module[78] and the quasi-harmonic analysis was done using the program CHARMM.[64] NAMD scripts to implement the three steps of the CSF calculation (i.e., confinement, hydration, and harmonic) are given in the Supporting Information.

The NAMD implementation of CSF was tested on the alanine dipeptide in explicit water. The confinement simulations (40 ns/run) were analyzed using the weighted-histogram confinement method, which yielded confinement free energies in quantitative agreement with those obtained in CHARMM; compare the results in Table S3 (Supporting Information) with those in Table S5 (Supporting Information). For the hydration free energy, FEP simulations in both forward and backward directions were performed using 10 windows and a linear scaling of λ, i.e., a window width of 0.1. The sampling time was 2.0 ns per window while saving configurations every 0.01 ps. Although the absolute hydration free energies computed by NAMD differ systematically by ∼0.3 kcal/mol from those obtained with CHARMM (Table S5, Supporting Information), this difference cancels out when the $\Delta\Delta F^{hydr}_{AB}$ is computed. Finally, the harmonic contribution to the conformational ΔF was obtained by substituting into eq 4 both the quasi-harmonic vibrational frequencies and the ground-state electronic energy, which was evaluated by subtracting the vibrational energy ($\sum_i^{\kappa} kT$) at the simulated temperature from the ensemble-averaged potential energy ($\langle V \rangle$) of the harmonically restrained conformers in a vacuum; see the Supporting Information for details. The QHA results are also given in Table S5 (Supporting Information). By introducing the confinement, the hydration, and the harmonic contributions into eq 6, conformational free-energy differences for all pairs of conformers of the alanine dipeptide were determined. The NAMD results are in quantitative agreement with those obtained by CHARMM; see Table 1.

Application to Biomolecules. The NAMD implementation of the CSF method (see above) was used to explore the conformational equilibria of more realistic biomolecules in solution, i.e., met-enkephalin (5 aa) and deca-alanine (10 aa). Once again, the CSF predictions were benchmarked against reference values obtained from converged molecular dynamics for met-enkephalin and literature calculations for deca-alanine.[25]

Met-Enkephalin. The difference in free energy between conformers I and IV of met-enkephalin (Figure 3A) was first estimated from a 3 μs simulation trajectory sampled at a constant temperature of 300 K. Given the complexity of the configurational space of this highly flexible peptide,[49] the free energy basins corresponding to these conformers were defined on the basis of the Cα-RMSD from reference structures. In particular, a Cα-RMSD cutoff of 0.75 Å was found to provide stable free energy results (Figure S9, Supporting Information). Using this definition, 219 round trips between conformers I and IV were collected along the trajectory and a reference value of $\Delta F = F_{IV} - F_I = -0.15 \pm 0.21$ kcal/mol was obtained. Although the large number of collected round trips should provide a very precise estimate of ΔF with a statistical error of 0.07 kcal/mol based on Figure 4, this is not what we observe from the bootstrapping analysis. In particular, using a definition of basins based on the principal-component analysis (PCA) of the interatomic distances,[79] only 22 round trips were identified yielding a $\Delta F_{I,IV}$ of 0.42 ± 0.22 kcal/mol, which points to systematic errors possibly arising from a crude (geometric) definition of the conformational basins. Because the aim of this report is to explore the performance of CSF rather than providing the most accurate estimate of $\Delta F_{I,IV}$ in solution, we remain with the definition based on the Cα-RMSD and compare the CSF predictions with results obtained from converged MD.

The equilibrium between the same conformational states of met-enkephalin was analyzed by CSF. To this aim and starting with energy-minimized structures, conformers I and IV were first thermalized at 300 K in the presence of the strongest restraint (k = 82 kcal/mol/Å^2). Then, confinement simulations of 20 ns per force constant were carried out in NAMD by decreasing the strength of the restraining potential stepwise. With the same definition of basins, i.e., a Cα-RMSD of 0.75 Å from the reference, a stable confinement free-energy difference of 4.21 ± 0.05 kcal/mol in favor of conformer I was obtained; see Table 2. Hydration free energies of the harmonically restrained conformers were then determined alchemically. Using the NAMD setup, i.e., a linear scaling of λ, 10 windows, 1 ns per window, and postprocessing via BAR, hydration free energies of −49.27 ± 0.67 and −51.69 ± 0.66 kcal/mol were obtained for conformers I and IV, respectively. Compared to results obtained with the alanine dipeptide, the standard error on the hydration free energy is at least 1 order of magnitude

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

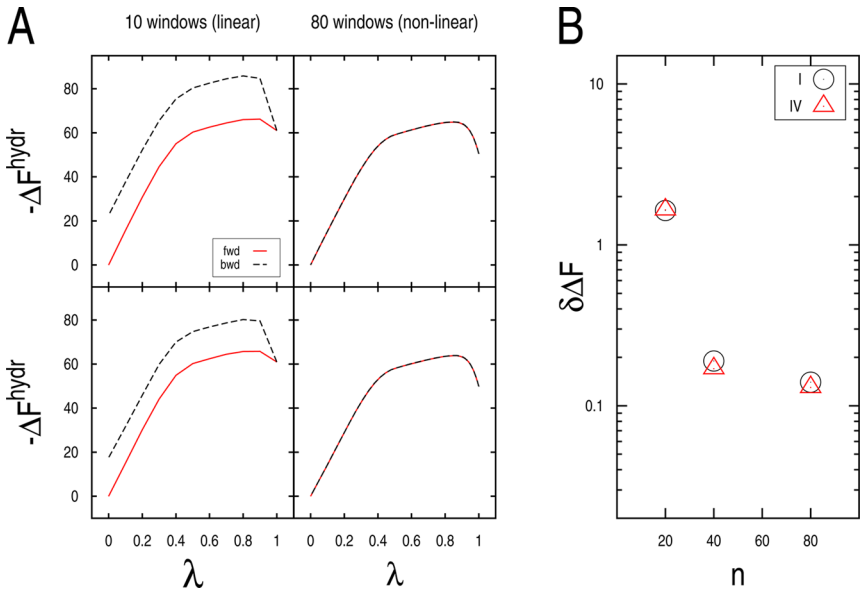


Figure 9. Errors in the hydration free energy calculation for the confined conformers of met-enkephalin. The hydration free energy was evaluated alchemically by decoupling the solute−solvent interactions in the forward direction (desolvation) and by recoupling them moving backward (hydration). (A) Systematic errors. The alchemical free energy along the forward and backward pathways is shown for conformers I (upper panel) and IV (lower panel) as determined by two FEP calculations implementing different stratification schemes: 10 windows with a linear scaling of λ (left) and 80 windows with a nonlinear scaling of λ (right), which is illustrated in Figure S9 (Supporting Information). The use of a denser stratification when the cavity hosting the solute forms/disappears together with a larger number of windows corrects for singularities arising from the repulsive term and ensures microscopic reversibility along the entire transformation. (B) Total error on the hydration free-energy calculations. BAR-estimated errors based on the nonlinear scaling of λ are shown as a function of the number of windows, n. Although similar hydration free energies are obtained with 20, 40, and 80 windows, the estimated error is greatly reduced (i.e., 1 order of magnitude) by increasing n from 20 to 80 for the same amount of sampling.

f9 larger. Moreover, as shown by Figure 9, there is a clear lack of microscopic reversibility in the forward versus backward transformations, which points to the occurrence of systematic errors. A detailed analysis of the hydration free-energy components for conformer I showed that during the alchemical transformation sizable errors may arise from the nonpolar contribution, which presents a large hysteresis (Figure S11, Supporting Information). In particular, because the largest deviations occur at the end of the desolvation path ($\lambda = 1$), i.e., when the cavity hosting the solute disappears in the forward direction or it forms moving backward, and are sensitive to the soft-core coupling scheme, we conclude that singularities in these calculations arise from the repulsive term. To correct for this, a denser and nonlinear stratification scheme was devised for the evaluation of the nonpolar contribution. Using the following scaling, i.e., $\lambda_i = 1 - (\delta_i - 1)^4$ with $\delta_i = i/n$ and $i = 0$, ..., n, and increasing the number of windows to 20, 40, or 80, which both provide denser stratifications at the end point, hydration free energies of -49.78 ± 0.08 and -50.27 ± 0.08 kcal/mol were obtained for conformers I and IV, respectively. As shown by Figure 9, the use of 80 windows with denser stratification at the end of the desolvation path ensures full microscopic reversibility and it reduces the total error on the hydration free energy by 1 order of magnitude. The CSF calculation was finally completed by a quasi-harmonic analysis of 40 ns trajectories of the two confined conformers in a vacuum, which yielded a harmonic free energy change of -3.88 ± 0.05 kcal/mol in favor of conformer IV.

Introducing the three contributions in eq 6, a $\Delta F_{I,IV}$ value of -0.16 ± 0.13 kcal/mol in favor of conformer IV was predicted by CSF, which is in quantitative agreement with the reference value obtained by ES, i.e., -0.15 ± 0.21 kcal/mol. Also, the CSF calculations smoothly converged to the correct free energy result with increasing k (Figure 10), which strongly supports f10 the new development. Finally, a component analysis of the CSF result (Table 2) indicates that the confinement contribution, which is large (4.21 kcal/mol) and favors conformer I, is canceled out by the hydration plus the harmonic contributions, which stabilize conformer IV. Interestingly, this observation suggests that the (marginally) higher thermodynamic stability of conformer IV in solution results from an enthalpy−entropy compensation with basin I stabilized entropically by its stronger anharmonicity and basin IV stabilized enthalpically by its more favorable electronic energy and interactions with the solvent, which plays a critical role in modulating the conformational equilibrium. This example nicely illustrates that the anharmonic contribution to the free energy of conformation of flexible molecules may be large, so that neglecting it would yield results off by several kcal/mol.

Deca-Alanine. Finally, the CSF method was applied to the α-helical (α) and β-hairpin (β) conformers of deca-alanine in solution; see Figure 3B. Because a reference value for ΔF could not be obtained from equilibrium sampling (no α to β transition could be sampled over μs-long MD simulations) or umbrella sampling (the absence of a "natural" reaction coordinate for this transition prevented convergence of the calculation), only a qualitative comparison with literature results[25] was possible. Using an all-atom RMSD cutoff of 2 Å to define the conformational basins consistent with ref 25, a $\Delta\Delta F^{conf}$ of 21.82 ± 0.09 kcal/mol in favor of the β-hairpin state was obtained. The hydration free energy of the confined states did not converge using the setup for met-enkephalin (20 windows) with BAR-estimated errors as large as 16.8 kcal/mol. Nonetheless, full convergence was achieved using denser

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

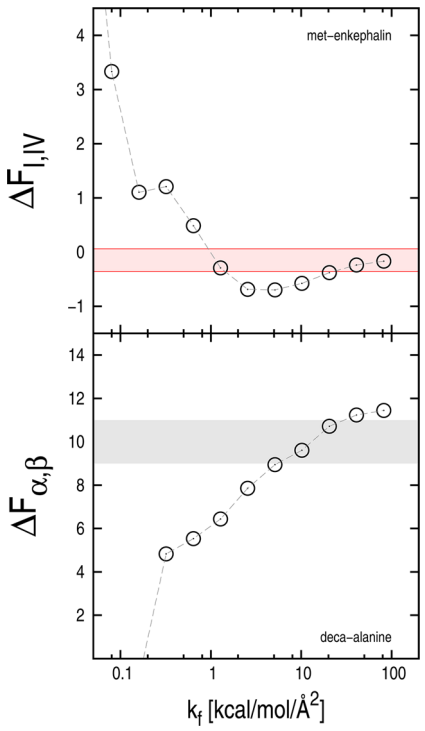


Figure 10. Convergence of the CSF results in the context of realistic biomolecules. On top, the difference in free energy between conformers I and IV of met-enkephalin is plotted as a function of the final confinement strength, k_f. The pink region corresponds to the reference ΔF (along with its error bar) obtained from equilibrium sampling. For k_f larger than 10 kcal/mol/Å^2, the CSF predictions smoothly converge to the correct result. On the bottom, the CSF results are plotted for the α to β transition of deca-alanine in explicit water. The gray region corresponds to the ΔF (along with its error bar) from literature calculations based on deactivated morphing.[25] As the reference structure for the β-hairpin was not strictly the same, the comparison with ref 25 should be considered as qualitative. At large k_f, the CSF results converge to a similar value of ΔF. All free energy values are given in kcal/mol.

nonlinear stratification schemes with 40 or 80 windows, which ensured microscopic reversibility along the entire alchemical transformation (Figure S12, Supporting Information). Also, the progressive reduction of the error bar for an increasing number of windows with no significant variation on the value of the free energy indicates that no systematic error was present. These denser stratifications provided hydration free energies for the α and β conformers of -60.12 ± 0.14 and -65.74 ± 0.25 kcal/mol, respectively. The apparent need for a large number of windows to obtain converged results suggests that the complexity of these calculations is clearly solute-size-dependent but that full control can be achieved through computationally more intensive stratifications. Finally, the CSF calculation was completed by a quasi-harmonic analysis, which yielded a $\Delta F_{\alpha^*\beta^*}$ value of 38.88 ± 0.06 kcal/mol in favor of the α conformer. By introducing all contributions into eq 6, CSF predicts a difference in free energy of 11.44 ± 0.31 kcal/mol in favor of the α-helical state in solution. This result is in qualitative agreement with the value of 10 ± 1 kcal/mol obtained by deactivated morphing[25] on structurally similar conformers. (Given the impossibility to obtain atomistic models for the α and β conformers from the authors of ref 25, the present comparison should be considered as qualitative.) We note that, because the difference in free energy between the α and β conformers of deca-alanine in solution is as large as 10 kcal/mol, the structural interconversion between them is slow and no spontaneous transition could be observed on the microsecond time scale. In addition, because the α/β transition involves a large structural change that requires the full reorganization of the backbone, umbrella sampling calculations over simple reaction coordinates like the end-to-end distance are inherently prone to fail[52] and never converged in our hands. In cases like the one illustrated here, the path-independent CSF method provides an elegant strategy to the conformational problem and should be preferred to both equilibrium MD and umbrella sampling.

CONCLUSION

The results presented in this paper indicate that the newly introduced confinement/solvation free energy (CSF) approach is a promising technique for the analysis of the free energy of conformation in biomolecules. The strong correlation with established free energy methods, the small error bars, and the smooth convergence of the calculations demonstrate that CSF is able to provide free energy estimates with chemical accuracy even for (small) biomolecules in a bath of explicit water. In this report, we have shown that (i) accurate conformational free energy differences can be obtained by CSF for folding peptides with an explicit treatment of the solvent; (ii) the convergence of the calculations is independent of the height of the barrier separating the conformers, which would dictate interconversion times at best of milliseconds for functional proteins; (iii) the computational strategy is path-independent, so that CSF is expected to yield accurate free energy results independently of the complexity of the conformational change and with no *a priori* knowledge of the transition pathway; and (iv) analysis of the CSF free-energy components provides fundamental insights on the conformational stability of biomolecules by separating contributions from the anharmonicity (entropy) and the solvent. Finally, a straightforward implementation of CSF into the highly scalable NAMD software has been presented. The latter will make the new development easily accessible and suitable for the CPU intensive exploration of conformational equilibria in larger biomolecules. Although the predictive power of the method needs to be further tested, the approach has the potential to become a useful tool for the theoretical investigation of protein function.

ASSOCIATED CONTENT

Supporting Information

The calculation of the hydration free energy; the determination of the harmonic free energy in a vacuum; the determination of conformational free energy differences from umbrella sampling calculations; analysis of both statistical and systematic errors in CSF; a free-energy component analysis for the alanine dipeptide, met-enkephalin, and deca-alanine; and a possible implementation of CSF into the highly scalable NAMD software. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mcecchini@unistra.fr.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

■ ACKNOWLEDGMENTS

We are grateful to Joel Montalvo Acosta for a critical reading of the manuscript. The work was granted access to the HPC resources of [CCRT/CINES/IDRIS] under the allocation 2014-[076644] made by GENCI (Grand Equipement National de Calcul Intensif). Access to computing resources of the MesoCentre at the University of Strasbourg is also acknowledged. J.E. received support from AFM Telethon (Grant No. 16141). Financial support from the International Center for Frontier Research in Chemistry (icFRC) is gratefully acknowledged. Efficient trajectory handling and part of the analysis were done with the help of the program Wordom.[80] This manuscript is dedicated to the memory of Francois Marchand for invaluable help in the early stages of the development.

■ REFERENCES

(1) Vale, R. D. Switches, Latches, and Amplifiers: Common Themes of G Proteins and Molecular Motors. *J. Cell Biol.* **1996**, *135*, 291−302.
(2) Palma, C.-A.; Cecchini, M.; Samorì, P. Predicting Self-assembly: from Empirism to Determinism. *Chem. Soc. Rev.* **2012**, *41*, 3713−3730.
(3) Monod, J.; Wyman, J.; Changeux, J. On the Nature of Allosteric Transitions: a Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88−118.
(4) Katz, B., Katz, B. *Nerve, Muscle, and Synapse*; McGraw-Hill: New York, 1966.
(5) Auerbach, A. The Energy and Work of a Ligand-gated Ion Channel. *J. Mol. Biol.* **2013**, *425*, 1461−1475.
(6) Chao, L. H.; Stratton, M. M.; Lee, I.; Rosenberg, O. S.; Levitz, J.; Mandell, D. J.; Kortemme, T.; Groves, J. T.; Schulman, H.; Kuriyan, J. A Mechanism for Tunable Autoinhibition in the Structure of a Human Ca^{2+}/Calmodulin-dependent Kinase II Holoenzyme. *Cell* **2011**, *146*, 732−745.
(7) Houdusse, A.; Sweeney, H. Myosin Motors: Missing Structures and Hidden Springs. *Curr. Opin. Struct. Biol.* **2001**, *11*, 182−194.
(8) Meirovitch, H. Recent Developments in Methodologies for Calculating the Entropy and Free Energy of Biological Systems by Computer Simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181−186.
(9) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(10) Kirkwood, J. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(11) Zwanzig, R. High-temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(12) Archontis, G.; Simonson, T.; Karplus, M. Binding Free Energies and Free Energy Components from Molecular Dynamics and Poisson-Boltzmann Calculations. Application to Amino Acid Recognition by Aspartyl-tRNA Synthetase. *J. Mol. Biol.* **2001**, *306*, 307−327.
(13) Zhou, R. Trp-cage: Folding Free Energy Landscape in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(14) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; et al. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91−97.
(15) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(16) Harvey, M.; Giupponi, G.; Fabritiis, G. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632−1639.
(17) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(18) Kumar, S.; Rosenberg, J.; Bouzida, D.; Swendsen, R.; Kollman, P. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(19) Roux, B. The Calculation of the Potential of Mean Force using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(20) Jorgensen, W. Free Energy Calculations: a Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(21) Zhuravlev, P.; Wu, S.; Potoyan, D.; Rubinstein, M.; Papoian, G. Computing Free Energies of Protein Conformations from Explicit Solvent Simulations. *Methods* **2010**, *52*, 115−121.
(22) Elber, R. Long-timescale Simulation Methods. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151−156.
(23) Zhu, F.; Hummer, G. Convergence and Error Estimation in Free Energy Calculations using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2012**, *33*, 453−465.
(24) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(25) Park, S.; Lau, A.; Roux, B. Computing Conformational Free Energy by Deactivated Morphing. *J. Chem. Phys.* **2008**, *129*, 134102.
(26) Freddolino, P.; Liu, F.; Gruebele, M.; Schulten, K. Ten-microsecond Molecular Dynamics Simulation of a Fast-folding WW Domain. *Biophys. J.* **2008**, *94*, L75.
(27) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the Free-energy Landscapes of Proteins by NMR-Guided Metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6817−6822.
(28) Meirovitch, H. Calculation of the Free Energy and the Entropy of Macromolecular Systems by Computer Simulation. *Rev. Comput. Chem.* **1998**, *12*, 1−74.
(29) Cheluvaraja, S.; Meirovitch, H. Simulation Method for Calculating the Entropy and Free Energy of Peptides and Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9241−9246.
(30) Tyka, M.; Clarke, A.; Sessions, R. An Efficient, Path-independent Method for Free-energy Calculations. *J. Phys. Chem. B* **2006**, *110*, 17212−17220.
(31) Cecchini, M.; Krivov, S.; Spichty, M.; Karplus, M. Calculation of Free-energy Differences by Confinement Simulations. Application to Peptide Conformers. *J. Phys. Chem. B* **2009**, *113*, 9728−9740.
(32) Terrell, L. *An Introduction to Statistical Thermodynamics*; Addison-Wesley Inc: Reading, MA. 1960; Chapter V.
(33) Frenkel, D.; Ladd, A. New Monte Carlo Method to Compute the Free Energy of Arbitrary Solids. Application to the fcc and hcp Phases of Hard Spheres. *J. Chem. Phys.* **1984**, *81*, 3188−3193.
(34) Stoessel, J. P.; Nowak, P. Absolute Free Energies in Biomolecular Systems. *Macromolecules* **1990**, *23*, 1961−1965.
(35) Ovchinnikov, V.; Cecchini, M.; Vanden-Eijnden, E.; Karplus, M. A Conformational Transition in the Myosin VI Converter Contributes to the Variable Step Size. *Biophys. J.* **2011**, *101*, 2436−2444.
(36) Roy, A.; Perez, A.; Dill, K.; MacCallum, J. Computing the Relative Stabilities and the Per-Residue Components in Protein Conformational Changes. *Structure* **2013**, *22*, 168−175.
(37) Ovchinnikov, V.; Cecchini, M.; Karplus, M. A Simplified Confinement Method for Calculating Absolute Free Energies and Free Energy and Entropy Differences. *J. Phys. Chem. B* **2012**, *117*, 750−762.
(38) Deng, Y.; Roux, B. Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.
(39) McQuarrie, D. *Statistical Mechanics*; Harper and Row: New York, 1976.
(40) Weeks, J.; Chandler, D.; Andersen, H. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(41) Karplus, M.; Kushick, J. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14*, 325−332.
(42) Maragakis, P.; Spichty, M.; Karplus, M. A Differential Fluctuation Theorem. *J. Phys. Chem. B* **2008**, *112*, 6168−6174.
(43) Ren, W.; Vanden-Eijnden, E.; Maragakis, P.; Weinan, E.; et al. Transition Pathways in Complex Systems: Application of the Finite-

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX

temperature String Method to the Alanine Dipeptide. *J. Chem. Phys.* **2005**, *123*, 134109.
(44) Gfeller, D.; de Los Rios, P.; Caflisch, A.; Rao, F. Complex Network Analysis of Free-energy Landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817−1822.
(45) Cruz, V.; Ramos, J.; Martínez-Salazar, J. Water-mediated Conformations of the Alanine Dipeptide as Revealed by Distributed Umbrella Sampling Simulations, Quantum Mechanics Based Calculations, and Experimental Data. *J. Phys. Chem. B* **2011**, *115*, 4880−4886.
(46) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. Probing the Competition between Secondary Structures and Local Preferences in Gas Phase Isolated Peptide Backbones. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1033−1048.
(47) Koneru, A.; Satyanarayana, S.; Rizwan, S. Endogenous Opioids: their Physiological Role and Receptors. *Global J. Pharmacol.* **2009**, *3*, 149−153.
(48) Sanbonmatsu, K.; Garcia, A. Structure of Met-enkephalin in Explicit Aqueous Solution using Replica Exchange Molecular Dynamics. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225−234.
(49) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. Exploring Multidimensional Free Energy Landscapes using Time-dependent Biases on Collective Variables. *J. Chem. Theory Comput.* **2010**, *6*, 35−47.
(50) Marcotte, I.; Separovic, F.; Auger, M.; Gagne, S. A Multidimensional 1H NMR Investigation of the Conformation of Methionine-enkephalin in Fast-tumbling Bicelles. *Biophys. J.* **2004**, *86*, 1587−1600.
(51) Hénin, J.; Chipot, C. Overcoming Free Energy Barriers using Unconstrained Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *121*, 2904−2914.
(52) Chipot, C.; Hénin, J. Exploring the Free-energy Landscape of a Short Peptide using an Average Force. *J. Chem. Phys.* **2005**, *123*, 244906.
(53) Kumar, S.; Payne, P.; Vasquez, M. Method for Free-energy Calculations using Iterative Techniques. *J. Comput. Chem.* **1996**, *17*, 1269−1275.
(54) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free Energy Calculation from Steered Molecular Dynamics Simulations using Jarzynski's Equality. *J. Chem. Phys.* **2003**, *119*, 3559−3566.
(55) Minh, D.; McCammon, J. Springs and Speeds in Free Energy Reconstruction from Irreversible Single-Molecule Pulling Experiments. *J. Phys. Chem. B* **2008**, *112*, 5892−5897.
(56) Forney, M. W.; Janosi, L.; Kosztin, I. Calculating Free-energy Profiles in Biomolecular Systems from Fast Nonequilibrium Processes. *Phys. Rev. E* **2008**, *78*, 051913.
(57) Hazel, A.; Chipot, C.; Gumbart, J. Thermodynamics of Deca-alanine Folding in Water. *J. Chem. Theory Comput.* **2014**, *10*, 2836−2844.
(58) Park, S., Khalili, F., Strumpfer, J. Stretching Deca-alanine Tutorial, http://www.ks.uiuc.edu/Training/Tutorials/science/10Ala-tutorial, 2012.
(59) MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(60) Mackerell, A.; Feig, M.; Brooks, C. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(61) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(62) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.
(63) Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(64) Brooks, B.; Brooks, C.; Mackerell, A.; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(65) Andersen, H. Molecular Dynamics Simulations at Constant Pressure and /or Temperature. *J. Phys. Chem.* **1980**, *72*, 2384−2393.
(66) Nose, S.; Klein, M. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(67) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-space Distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(68) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(69) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1989.
(70) Shirts, M. R.; Mobley, D. L. An Introduction to Best Practices in Free Energy Calculations. *Biomolecular Simulations*; Springer: New York, 2013; pp 271−311.
(71) Krivov, S.; Muff, S.; Caflisch, A.; Karplus, M. One-dimensional Barrier-preserving Free-energy Projections of a β-sheet Miniprotein: New Insights into the Folding Process. *J. Phys. Chem. B* **2008**, *112*, 8701−8714.
(72) Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo Methods describe Noncovalent Interactions with Subchemical Accuracy. *J. Chem. Theory Comput.* **2013**, *9*, 4287−4292.
(73) DeLano, W. L. *PyMOL Molecular Graphics System*, https://www.pymol.org//, 2002.
(74) Hub, J.; de Groot, B.; van der Spoel, D. g_wham - A free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(75) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289−6292.
(76) Hénin, J., Gumbart, J., Harrison, C., Chipot, C. Free Energy Calculations along a Reaction Coordinate: A Tutorial for Adaptive Biasing Force Simulations, http://www.ks.uiuc.edu/Training/Tutorials/namd/ABF/tutorial-abf.pdf, 2012.
(77) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. A Toolkit for the Analysis of Free-energy Perturbation Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2606−2616.
(78) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111*, 3345−3362.
(79) Elmaci, N.; Berry, R. Principal Coordinate Analysis on a Protein Model. *J. Chem. Phys.* **1999**, *110*, 10606−10622.
(80) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007**, *23*, 2625−2627.

DOI: 10.1021/acs.jpcb.5b01632
J. Phys. Chem. B XXXX, XXX, XXX−XXX