

Force Field Bias in Protein Folding Simulations

Peter L. Freddolino,[†] Sanghyun Park,[‡] Benoît Roux,[§] and Klaus Schulten^{†*}

[†]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois; [‡]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois; and [§]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois

ABSTRACT Long timescale ($>1 \mu\text{s}$) molecular dynamics simulations of protein folding offer a powerful tool for understanding the atomic-scale interactions that determine a protein's folding pathway and stabilize its native state. Unfortunately, when the simulated protein fails to fold, it is often unclear whether the failure is due to a deficiency in the underlying force fields or simply a lack of sufficient simulation time. We examine one such case, the human Pin1 WW domain, using the recently developed deactivated morphing method to calculate free energy differences between misfolded and folded states. We find that the force field we used favors the misfolded states, explaining the failure of the folding simulations. Possible further applications of deactivated morphing and implications for force field development are discussed.

INTRODUCTION

Computational efforts to study the folding process of proteins have long accompanied experimental studies of protein folding pathways and kinetics, beginning with lattice models (reviewed in (1)) and more recently progressing to atomistic or coarse-grained simulations of protein folding trajectories or folding-unfolding equilibria (e.g., (2–6)). Such simulations have always faced multiple technical and practical challenges, chief among them problems of timescale and accuracy. The fastest folding full-length proteins currently known require 0.7–1.0 μs to fold (7,8) with a hypothesized limit at approximately $(N/100) \mu\text{s}$ for an N residue protein (9); only recently have atomistic molecular dynamics (MD) simulations with explicit solvent on microsecond timescales become feasible (10–12). At the same time, all MD simulations are reliant on the force field used to both correctly identify the native state of the protein as lowest in free energy and to provide a realistic description of the intermediate structures encountered during folding and the transitions between them (13). While numerous successes of all-atom MD in folding proteins to near-native or native states illustrate that in many cases an accurate treatment is possible, recent studies have also found hurdles such as a preference for helical structures in many force fields (14–16) and failures to consistently rank folded structures of proteins as free energy minima (17,18). Additional concerns have been noted for implicit-solvent folding simulations, which may yield free energies for folding intermediates that are different from those obtained through explicit solvent MD (18,19).

The combination of recent research into identifying fast folding proteins (8,20,21), improving performance of molecular dynamics software, and ever-increasing scientific

computing resources has made it possible to perform the multiple microsecond, explicit solvent MD simulations necessary to observe complete protein folding events. Using a selected subset of the machines available to their Folding-home effort, Ensign et al. recently presented a large ensemble of folding trajectories for a fast folding villin mutant (11). We recently reported a 10- μs trajectory of the fast-folding human Pin1 WW domain mutant Fip35 beginning from an unfolded state; however, an array of helical structures was observed instead of the expected β -sheet structure (12).

WW domains are small, antiparallel three-strand β -sheet proteins; we focus on the human Pin1 WW domain, which has a well-characterized folding mechanism in which formation of the first turn is the rate-limiting step (22). More recent experimental studies on a variety of Pin1 WW domain mutants showed that mutants with increased melting temperatures tended toward incipient downhill or downhill folding behavior (23). Simulations of Pin1 WW domain folding using Gō-like models have yielded a variety of hypotheses for the order of folding, either with the hydrophobic core forming first (24) or last (25), and with the potential for different orders of β -sheet assembly at different temperatures (26).

All-atom molecular dynamics simulations using physics-based potentials could provide valuable additional information on the folding process of the Pin1 WW domain and aid in distinguishing between proposals from previous simulations, allowing for greater understanding of the role of factors such as nonnative interactions and solvent effects. However, the failure of a 10- μs trajectory starting from a fully denatured Pin1 WW domain mutant to show any progress toward a natively like structure (12) raises the specter of the challenges to molecular dynamics discussed earlier. Clearly the simulation failed to properly treat WW domain folding, but whether this failure is due to kinetic trapping, a fundamental thermodynamic problem in the force field, or some other issue, is unclear. To distinguish between these possibilities, we performed three additional multiple-microsecond

Submitted December 4, 2008, and accepted for publication February 19, 2009.

*Correspondence: kschulte@ks.uiuc.edu

Editor: Angel E. Garcia.

© 2009 by the Biophysical Society
0006-3495/09/05/3772/9 \$2.00

doi: 10.1016/j.bpj.2009.02.033

folding simulations of the WW domain from different starting conditions. In addition, the recently developed deactivated morphing (DM) method (27) was used to calculate free energy differences between the native state and three commonly observed helical states in the folding trajectories. We find that under the simulation conditions used, all three helical states are favored over the native state by 4.4–8.1 kcal/mol. In addition, with defined free-energy differences between chosen reference structures, the effects of perturbations to the bonded and nonbonded parameters on the overall free energy differences between conformations can be studied. The DM procedure can thus be used in establishing a foundation for other free energy calculations where more than one conformational state of a protein must be studied, and in testing the effects of alterations to the potential energy functions used for MD simulations.

METHODS

Molecular dynamics

All simulations were carried out using the development version of NAMD 2.7 (28). As in Freddolino et al. (12), appropriate mutations were applied to a Pin1 WW domain crystal structure (PDB code 2F21 (29)) to yield the sequence described in variant 23 from Liu et al. (23), which is referred to as Fip35. The protein was solvated in a cubic box of 10,014 TIP3P water molecules and neutralized with 30 mM NaCl using VMD (30). Starting structures for folding simulations were generated either by setting all (ϕ , ψ) angle pairs to $(-135, 135)$ for SimFold₁ and SimFold₂, or for SimFold₃ and SimFold₄ by generating two separate thermally denatured states through simulation at 490 K for 100 ns, in both cases yielding structures with no sheet or helix structure (as calculated by STRIDE (31)). The denatured starting structures for SimFold₃ (SimFold₄) had Q_{res} (32) of 0.139 (0.183), C_{α} -RMSDs to the folded structure of 13.1 Å (10.2 Å), and radii of gyration of 14.0 Å (12.8 Å) (compare to 9.7 in the folded structure), respectively. The starting structures were subjected to 3000 steps of minimization and 100 ps of NVT equilibration before production runs, using a periodic cell size obtained from a 100-ps NPT equilibration of the wild-type Pin1 WW domain structure. For SimFold₁ and SimFold₂ different initial velocities (and a different series of random number seeds for the thermostat) were used, although the initial protein conformations were identical.

Except where otherwise noted, the CHARMM22 force field with CMAP corrections (33) was used for the protein. Short-range nonbonded interactions were cut off at 8.0 Å with switching beginning at 7.0 Å; long-range electrostatics was treated using the particle-mesh Ewald method. All bonds involving hydrogens in the protein were constrained using the RATTLE algorithm (34) with water geometry maintained using SETTLE (35). An integration timestep of 2.0 fs was used, with bonded and short-range interactions evaluated every timestep and long-range electrostatics once every three timesteps. A temperature of 337 K was maintained using a Langevin thermostat with a damping constant of 0.1 ps⁻¹. Cluster analysis used the *g_cluster* module of GROMACS 3.3 (36) with the GROMOS clustering method (37); further details are provided in Cluster Analysis of Folding Trajectories in Supporting Material. Free energy calculations (see below) were performed in the NPT ensemble using a Nosé-Hoover Langevin piston barostat (28) with a period of 200.0 fs and damping timescale of 100.0 fs. For the free energy calculations, a Langevin damping constant of 1.0 fs⁻¹ was used, and coordinates and other data were saved once every 100 fs.

Free energy calculations

Deactivated morphing (DM) between sheet and each of HELIXU, HELIXL, and HELIXV was performed as described in Park et al. (27), with the details of all

calculations and modifications of the original DM procedure described in The Deactivated Morphing Process in Supporting Material. In brief, the calculation of the conformational free energy difference between any two reference conformations is divided into a series of steps between intermediates. We refer to the unrestrained ensemble of structures within a specified protein RMSD cutoff of a reference conformation as E ; the state with harmonic restraints applied to all protein atoms restraining it to the reference conformation with $\kappa = 1000$ kcal/mol Å as K ; the deactivated state with all protein atoms restrained to their coordinates in reference state as Q ; and a “dummy” state with a uniform set of van der Waals parameters and charges applied as D (see The Deactivated Morphing Process in Supporting Material). Calculation of the free energy difference between the unrestrained ensembles $E(A)$ and $E(B)$ for reference conformations A and B is thus performed by following a path from $E(A)$ through the increasingly restrained states to $D(A)$, then morphing $D(A)$ to $D(B)$ along the least-squares path (38), and finally following a path of decreasing restraints to $E(B)$; this process is shown schematically (see Fig. 3). Each of the transitions is further subdivided to provide sufficient overlaps between adjacent states, as detailed in the Supporting Material. For the case of morphs involving HELIXU or HELIXV, an additional step needed to be taken to account for the effects of a site-bound water in the reference structure (see Site-bound waters, in The Deactivated Morphing Process, Supporting Material); the free energy difference from this added step is included in the morphing step (see Fig. 3) and in the discussion below.

Error analysis for all free energy calculations was performed using block averaging; all data were split into 10 blocks, the first block was discarded, and then free energy calculations were performed independently for the data in each of the nine remaining blocks, with the mean of these block estimates corresponding to the reported value and error bars given as $\pm 2\sigma/\sqrt{9}$, with σ the standard deviation of the block estimates.

RESULTS

To aid in ruling out kinetic trapping or a single pathological trajectory in the failure of WW domain folding in previous simulations, three long-timescale folding trajectories were run, one (SimFold₂) starting from the same fully extended conformation as in the previous work (12) and run for 3.4 μ s, and two (SimFold₃ and SimFold₄) starting from heat-denatured structures (see Materials and Methods) and run for 4.1 and 4.4 μ s, respectively. For reference, the 200-ns native state simulation and 10.0- μ s folding trajectory from our previous work are denoted SimCryst and SimFold₁, respectively.

The secondary structure, fraction of native contacts Q_{res} (32), exposed hydrophobic surface area, and C_{α} -RMSD observed in SimFold₂, SimFold₃, and SimFold₄ are plotted in Fig. 1. A similar pattern to SimFold₁ is observed in SimFold₂ and SimFold₃; the trajectories show rapid hydrophobic collapse, followed by the formation of mostly helical structure, which persists throughout the simulation. SimFold₄ takes longer to reach helical conformations, as it spends most of the duration of the trajectory after its initial collapse in a coil state stabilized by an intricate network of salt bridges; however, even in this trajectory, stable helical structure forms after 2.5 μ s. In no case is significant progress toward the native state observed, nor is any persistent β -structure formed. For reference, as reported previously, SimCryst showed that Fip35 is remarkably stable in a crystal-structure-like conformation over the 200-ns

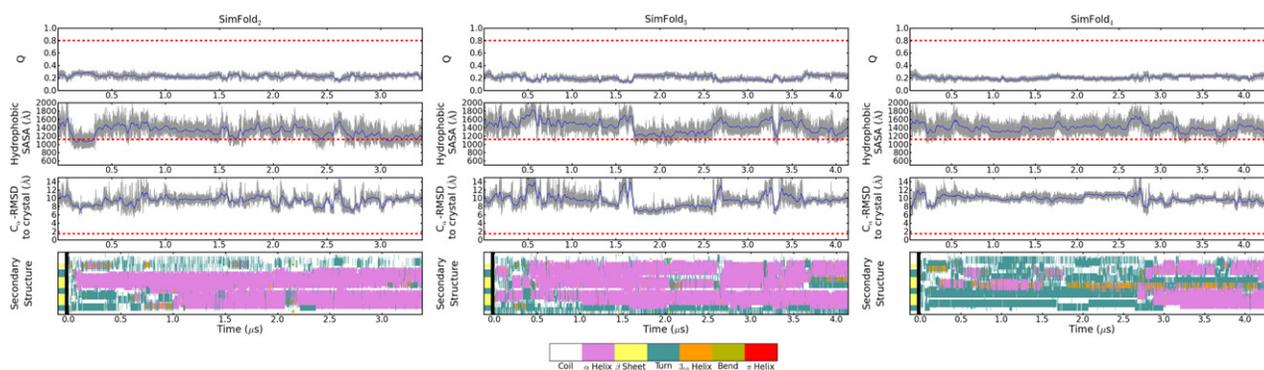


FIGURE 1 Properties of the WW domain in simulations SimFold₂, SimFold₃, and SimFold₄. For Q_{res} , solvent-accessible surface area of hydrophobic groups, and C_{α} -RMSD to the crystal structure, mean values from SimCryst for the native state are shown as dashed lines. Secondary structure throughout the simulations is plotted using the color scale shown below the figure.

simulation, with no structures showing a C_{α} -RMSD >2.0 Å, and only one frame (at ~ 61 ns) with an all-protein atom RMSD >4.0 Å, relative to the crystal structure.

As was the case for SimFold₁, clustering analysis (see Cluster Analysis of Folding Trajectories, [Supporting Material](#)) illustrates that there is no one stable helical conformation being formed in any of the new folding trajectories; instead, a series of interconverting helical states are observed. While suggesting that the previous finding of mostly helical structure in SimFold₁ was not a statistical anomaly, however, these additional simulations still cannot rule out the possibility that what is observed in the WW domain simulations is simply a case of conformational trapping, and that given a sufficiently long simulation (beyond the experimental folding timescale of the WW domain), the proper β -sheet structure would form.

To decisively distinguish between kinetic and thermodynamic problems in WW domain folding simulations, deactivated morphing calculations (27) were performed to calculate conformational free energy differences between the folded state and a set of three commonly occurring helical structures observed in the folding simulations.

Selection of reference structures

Deactivated morphing requires both the definition of reference structures for all endpoints being used, and choice of a criterion to define the conformational ensemble which will be included as part of that endpoint. Here we use the all-atom RMSD relative to the pertinent reference state to define the conformational ensembles; based on the observed fluctuations in SimCryst, a 4.0 Å cutoff is assumed unless otherwise noted.

In the absence of any obvious reference structure for the misfolded helical states, clustering analysis was performed using a 4.0 Å all-atom RMSD cutoff on each of SimFold₁, SimFold₂, and SimFold₃ to identify candidate reference structures. For each identified cluster, the structure from that cluster with the lowest pairwise RMSD to all structures in the same cluster was taken as a representative, and the number of struc-

tures in SimFold₁, SimFold₂, and SimFold₃ within 4.0 Å of this reference structure was calculated. For each simulation, the structure that included the largest total number of frames from the folding simulations in its defined conformational ensemble was taken as a helical reference conformation and further refined (see The Deactivated Morphing Process in the [Supporting Material](#)) to yield HELIXU (from SimFold₁), HELIXV (from SimFold₂), and HELIXL (from SimFold₃); in the case of SimFold₁ the second-highest occupancy cluster was used since the highest occupancy cluster was nearly identical to HELIXL. These reference conformations, along with the sheet state from SimCryst, are shown in Fig. 2. The fraction of frames from each folding simulation within 4.0 Å of the DM reference states is shown in Table 1; we note that

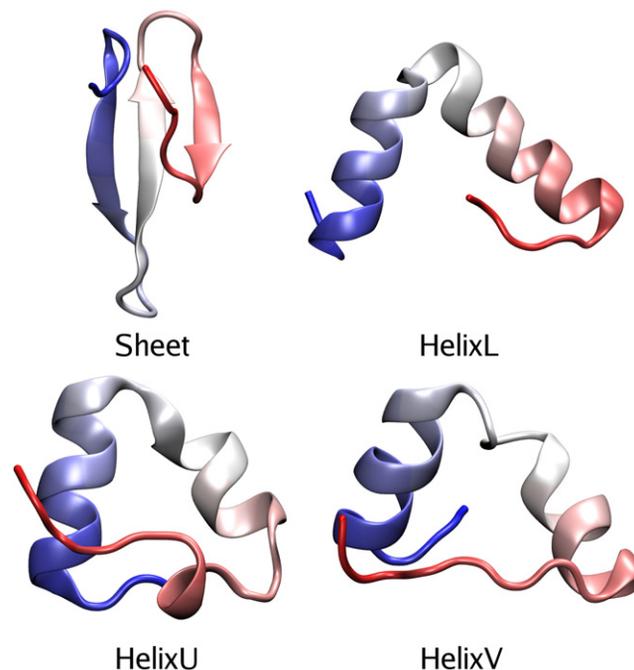


FIGURE 2 Cartoon representations of the reference structures used as endpoints for deactivated morphing simulations. Color scale runs blue to red from N-terminus to C-terminus.

TABLE 1 Durations of all trajectories in this study, and the fraction of timesteps from each trajectory within 4.0 Å of each of the DM reference states

Simulation	Duration (ns)	Fraction of timesteps within 4.0 Å of			
		SHEET	HELIXL	HELIXU	HELIXV
SimCryst	(200)	1.000	0.000	0.000	0.000
SimFold ₁	(10000)	0.000	0.053	0.042	0.025
SimFold ₂	3384	0.000	0.000	0.014	0.048
SimFold ₃	4139	0.000	0.068	0.000	0.000
SimFold ₄	4357	0.000	0.000	0.000	0.000

Durations in parentheses indicate trajectories that were originally reported in Freddolino et al. (12). The total simulation time shown is 22 μ s.

while these populations may seem rather small, no more inclusive set of reference states could be identified, a fact attributable to the wide variety of different helical conformations observed in the folding simulations. The predominant helical structure formed during the latter half of SimFold₄ is distinct from the common helical structures in the other three trajectories, and is not considered in the deactivated morphing calculations. While the use of other distance metrics such as dRMSD (39,40) for clustering might allow the identification of more consolidated clusters, RMSDs were used to maintain compatibility with the definitions of end states for deactivated morphing. It must be noted that SimFold₃, and thus the structure chosen for HELIXL, contains a *cis*-peptide bond at PRO3. As seen in the clustering results, the structures occurring in SimFold₃ were very similar to those in SimFold₁ and SimFold₂, likely because the N-terminus is part of an N-terminal coil in all major conformations observed, but the DM results for HELIXL sample only conformations with a *cis* bond at residue 3, and it is not clear exactly how similar the free energy of HELIXL-like conformations (such as those observed in SimFold₁ and SimFold₂) with a *trans* bond at PRO3 would be.

Conformational free energy differences between states

A schematic showing the free energy differences associated with each DM step is given in Fig. 3, with all pairwise free energy differences for the four reference structures shown in Table 2. All three helical states are significantly lower in free energy than the native state, and the free energy gap is observed to widen for larger ensemble cutoffs, likely reflecting the significant conformational freedom observed in the misfolded helical states during folding simulations. Given the presence of at least three local free energy minima corresponding to helical states which are several $k_B T$ below the free energy of the native state, it is not surprising that the WW domain did not fold in our simulations, and even longer simulations using the same parameters should not significantly occupy the folded state. The *cis*-proline containing HELIXL conformation is notably lower in free energy than any of the all-*trans* cases; it appears that this conformation is stabilized by the *cis*-PRO3, since trajectory SimFold₃ showed little population of HELIXU and HELIXV-like conformations, whereas SimFold₁ and SimFold₂ show more even population of conformations resembling the DM reference structures (likely indicating that HELIXL-like conformations with *trans*-PRO3 are similar to HELIXU and HELIXV in stability).

Decomposition of the conformational free energy difference into the changes associated with different DM steps (see Methods) can yield qualitative insight into factors contributing to the observed free energy difference. To a first approximation, the restraint step ($E \rightarrow K_1$) primarily represents the conformational and vibrational entropy of the protein in a given conformational well, the deactivation step ($K_1 \rightarrow Q$) represents the internal enthalpic interactions of the reference structure, and the combination of dummyming

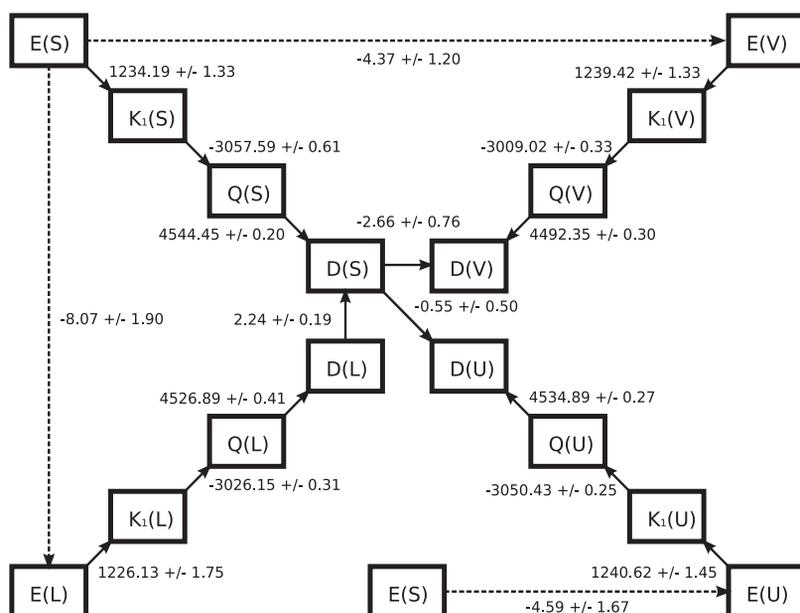


FIGURE 3 Schematic of deactivated morphing between SHEET (S) and three helical conformations (L, U, and V for HELIXL, HELIXU, and HELIXV). Solid arrows represent transitions, which were calculated using deactivated morphing, with the dashed arrows showing the sum over the path of calculated transitions between two unrestrained states. All energies given in kcal/mol.

TABLE 2 Conformational free energy changes calculated via deactivated morphing for a transition from the state shown on top to the state shown at left

State-state free energy differences (kcal/mol)				
State	SHEET	HELIXL	HELIXU	HELIXV
SHEET				
2.0	—	7.28 ± 1.84	1.37 ± 1.65	0.08 ± 1.59
3.0	—	7.99 ± 1.84	3.75 ± 1.64	2.53 ± 1.37
4.0	—	8.07 ± 1.90	4.59 ± 1.67	4.37 ± 1.20
HELIXL				
2.0	-7.28 ± 1.84	—	-5.76 ± 1.73	-6.76 ± 1.26
3.0	-7.99 ± 1.84	—	-4.08 ± 2.00	-5.02 ± 1.13
4.0	-8.07 ± 1.90	—	-3.33 ± 2.16	-3.26 ± 1.16
HELIXU				
2.0	-1.37 ± 1.65	5.76 ± 1.73	—	-1.65 ± 1.26
3.0	-3.75 ± 1.64	4.08 ± 2.00	—	-1.59 ± 1.18
4.0	-4.59 ± 1.67	3.33 ± 2.16	—	-0.58 ± 1.31
HELIXV				
2.0	-0.08 ± 1.59	6.76 ± 1.25	1.65 ± 1.26	—
3.0	-2.53 ± 1.37	5.02 ± 1.13	1.59 ± 1.18	—
4.0	-4.37 ± 1.20	3.26 ± 1.16	0.58 ± 1.31	—

The numbering included with the destination states refers to the cutoff (in Å) used to define the unrestrained ensembles at both endpoints.

(Q → D) and morphing (D → D) steps yields the difference in interactions with water between the two structures. The calculated free energy differences may be further understood through comparison with a number of characteristics of the reference states shown in Table 3.

Based on the rough breakdown of free energy contributions above, the free energies from the restraining step suggest that HELIXU and HELIXV have similar levels of conformational entropy, both greater than SHEET, and HELIXL is significantly more constrained than three of the other states. While one might expect, based on AFM data, that all three helical states would be lower in conformational entropy than the sheet (41), both HELIXU and HELIXV have a relatively free C-terminal coil that likely contributes to their conformational freedom; it should also be noted that the free energy differences between these structures and sheet drops significantly for lower RMSD cutoffs defining the unrestrained ensemble $E(A)$ (see Methods). Given the increasing favor-

TABLE 3 Energetic and qualitative contributions for the DM reference conformations; ΔG_{solv} is the Poisson-Boltzmann solvation free energy

	Properties of DM reference conformations			
	SHEET	HELIXL	HELIXU	HELIXV
$U_{\text{P, short}}$ (kcal/mol)	2902.05	2870.45	2893.96	2852.60
ΔG_{solv} (kcal/mol)	-403.82	-526.58	-350.23	-268.28
Hydrogen bonds	21	27	21	24
Backbone H-bonds	11	19	13	13
Salt bridges	3	3	5	6
SASA (Å ²)	2972.7	3071.1	3143.7	3019.7
Hydrophobic SASA (Å ²)	1071.7	1301.2	1137.2	1319.5

Hydrogen bonds were calculated using a 3.5 Å heavy atom distance cutoff and 35° donor-hydrogen-acceptor angle cutoff.

ability of HELIXU and HELIXV throughout the range of RMSD cutoffs that we included, it is likely that the helical states would be even more heavily favored for cutoffs above 4.0 Å, due to the wide variety of accessible helical conformations present in the trajectories; however, obtaining converged results for the restraining step becomes increasingly more difficult for larger cutoff values, and thus we restrict ourselves to a maximum of 4.0 Å all-atom RMSD.

Based on the energies from the deactivation step, all three helical states appear to have more favorable protein-protein interactions than SHEET, which is also apparent in the total energy of internal interactions, $U_{\text{P, short}}$ (defined as all short-range interactions between protein atoms). Inspection of polar interactions in the reference states (see Table 3) illustrates that two of the three helical states contain more protein-protein hydrogen bonds than sheet, and that all three contain more backbone-backbone hydrogen bonds. In addition, HELIXU has two more protein-protein salt bridges, and HELIXV three more than SHEET and HELIXL.

One particular concern that has been commonly raised in recent discussions of classical MD force fields is the treatment of hydrogen bonding primarily as a dipole-dipole interaction, which is known in some cases to lead to deviation from experimental and high-level theory data (42,43) and, for proteins, from the expected distribution of geometries from high-resolution crystal structures (44). To analyze the protein-protein hydrogen bonding geometries of the conformations in this study, the hydrogen-acceptor distance δ_{HA} , angle at the hydrogen Θ , angle at the acceptor Ψ , and acceptor antecedent dihedral angle X were plotted for a set of frames from the most weakly restrained states of the DM calculations (here we use the notation and definitions from (45)). These data are shown in Hydrogen Bonding Analysis in the Supplementary Material. The distributions of δ_{HA} , Θ , and X for backbone-backbone and sidechain-sidechain hydrogen bonds follow the general patterns observed in a survey of x-ray crystal structures (in (45)). In the case of Ψ angles, the backbone hydrogen bonds of the helical state follow the expected distribution, with a maximum occurring at $\sim 155^\circ$, likely due to other constraints imposed by this secondary structure (45); however, the distribution for backbone hydrogen bonds in SHEET shows that the backbone-backbone hydrogen bonds here are mostly linear, and in both HELIXL and HELIXV the sidechain-sidechain hydrogen bonding also showed overpopulation of near-linear conformations. It is, however, somewhat remarkable that the Ψ angle distributions for side-chain hydrogen bonds were in all cases peaked near 120° rather than 180° .

The combined free energies for dummied and morphing steps (i.e., the path $Q(S) \rightarrow D(S) \rightarrow D(H) \rightarrow Q(H)$) for going from SHEET to the helical states HELIXL, HELIXU, and HELIXV yield ΔG values of 15.32, 9.01, and 49.44 kcal/mol, respectively, suggesting that water interacts more favorably with sheet than any of the helical states. It should be noted that the favorability of sheet likely comes from specific polar

interactions with water, since $D(S)$ is less stable than the dummied helical conformations. As seen in Table 3, both the overall solvent-exposed surface area (SASA) and hydrophobic SASA are lower for sheet than the helical states. Of these, the overall protein SASA is expected to be approximately proportional to the entropy of solvation for a large hydrophobic or heterogeneous solute, although corrections are required for factors such as excluded volume, attractive solvent-solute interactions, and deviations from ideal behavior due to exposure of portions of solute too small for the SASA-based approximation to hold (46–50); the hydrophobic SASA simply provides a qualitative measure of how well packed the conformations are and how favorable the total free energy of solvation may be.

Effects of changes in the potential

The presence of calculated conformational free energy differences between the native state and multiple nonnative states that are favored by the force field allows the testing of changes to the potential which might be expected to change the ranking of these states, simply by calculating the effect of a proposed change on the free energy of each of the structural ensembles used in DM. Given that a relatively short real-space nonbonded interaction cutoff of 8 Å was used in the folding simulations, one can determine whether using a larger cutoff of 12 Å would stabilize the native state. Inspired by previous calculations on a WW domain in which backbone electrostatics and torsional terms were investigated to determine their effects on structure (14), we also calculated the internal energy change $\Delta U_{P,short}$ of the protein in each reference conformation in the presence of increased and decreased backbone polarization, and with the CMAP correction removed (i.e., using the CHARMM22 backbone potential), in an effort to identify modifications that might favor the native state (data not shown). Only the removal of CMAP favored SHEET relative to all three helical states, and thus the free energy change associated with removing CMAP corrections was also calculated for all four conformations. The complete set of intermediates simulated to investigate the effects of these changes to the potential is shown in Fig. S13 in the Supporting Material, with the calculated energies shown in Table 4.

As expected given the use of long-range electrostatics, expanding the short-range nonbonded cutoff from 8.0 Å to 12.0 Å does not significantly alter the free energy differences between states. Likewise, the removal of CMAP corrections

does not stabilize sheet relative to HELIXU or HELIXV, and actually makes it less favorable relative to HELIXL. Thus, neither of the simple perturbations to the potential tested here significantly alters the helix-sheet equilibrium from the values obtained via deactivated morphing; the DM results do, however, provide a convenient scaffold for testing other modifications of the potential through a similar mechanism without needing to redo the complete DM calculations. It must be noted that the nonnative reference states chosen using the original potential will not necessarily represent local free energy minima in the new potential, and thus for some applications (for example, if alterations to the potential caused the native state to become more favorable than the misfolded states in this study) it may be necessary to perform further conformational sampling in the new potential to identify appropriate reference states.

DISCUSSION

After consistent failure of the WW domain to fold into a native or near-native state over microsecond timescales in silico, one was left with the question of whether this failure was due to kinetic or thermodynamic inaccuracies. The deactivated morphing calculations presented here firmly indicate the latter; three helical misfolded states observed in folding simulations were all found to be $>6 k_B T$ lower in free energy than the native state. A set of related questions now arises; among them, questions of what energetic factors lead to this free energy gap, how the force field and simulation conditions might be modified to yield proper relative stabilities for different conformations of the WW domain, and whether the failure for this protein reflects a more systematic problem that would also occur for other, similar proteins.

A number of recent studies have indicated that modern MD force fields, including the CHARMM force field, slightly overestimate the presence of helical structure in small peptides (15,51). This may be associated with systematic errors in the protein backbone potential (investigated in (52)) along with inherent limitations in the simplified energy function used in protein force fields. Comparison of backbone-backbone hydrogen bonding patterns in the DM reference states to data from a survey of crystallographic structures (45) illustrated irregularities in the distribution of angles centered on the hydrogen and acceptor atoms, which might significantly affect the relative stability of the studied structures. Further calculations using a potential with explicit hydrogen bonding terms or other features designed to refine the treatment of

TABLE 4 Free energy changes for perturbing the potential in each of the four reference states

	Free energy differences for perturbations (kcal/mol)			
	SHEET	HELIXL	HELIXU	HELIXV
$A \rightarrow B$	28.79 ± 0.84	24.58 ± 1.70	27.23 ± 1.86	29.01 ± 1.51
$A \rightarrow C$	-716.49 ± 0.48	-716.37 ± 0.51	-717.04 ± 0.63	-716.43 ± 0.82

Endpoints are shown in Fig. S13; A, B, and C refer to states with the original simulation parameters, no CMAP correction, and 12 Å cutoffs, respectively.

hydrogen bonds (such as inclusion of virtual sites (53)) would be required to assess the effects of improved descriptions of hydrogen bonding on the structure of Fip35. The helical states of the WW domain in our simulations are also likely stabilized by the presence of additional protein-protein hydrogen bonds and salt bridges in all cases. Any inaccuracy in backbone hydrogen-bonding treatment, such as those discussed above, could also extend to sidechain-sidechain and sidechain-backbone hydrogen bonds and salt bridges.

Another possible cause for the overstabilization of helical states observed here is an artifact related to the system's periodicity. In a series of comparisons between MD simulations using Ewald electrostatics and continuum electrostatics calculations, it was found that unduly small periodic cells could overstabilize helical conformations of short polypeptides (54) or other compact conformations (55), and lead to biases on the order of a few $k_B T$ between different conformations in a nanosecond timescale MD trajectory of a 66-residue protein (56). While such periodicity artifacts cannot be ruled out in this study, a number of factors suggest that any such effects do not, on their own, lead to the free energy differences observed between WW domain conformations: the WW domain is smaller and less heavily charged than the solutes characterized in Kastenholz and Hunenberger (56), and is simulated in dilute NaCl in a large periodic cell (6.8 nm cube) given the size of the solute relative to the simulations in that study (56). In addition, the reference conformations being considered are similarly compact relative to each other, and the free energy differences between states are significantly larger than the maximum biases reported in Kastenholz and Hunenberger (56).

While the calculations presented here cannot firmly establish what factor or combination of factors in the force field leads to overstabilization of helical conformations of the WW domain, the decomposition of the free energy differences into differences between a series of intermediates indicates that internal protein-protein interactions favor the helical states, interactions with water (polar and apolar) favor the sheet state, and that HELIXU and HELIXV both have significant contributions to their conformational entropy from the wide variety of accessible conformations near the reference state. These contributors, of course, cannot be fully separated, as (for example) a refinement of the potential governing protein-protein interactions would also significantly affect the conformational entropy of the different states. The overpopulation of near-linear hydrogen bonding geometries in several of the simulations suggests one possible point of improvement for the potential, which could also significantly affect the overall helix/sheet ratio observed using CHARMM22 and other modern force fields. Sources of error other than the treatment of hydrogen bonding are of course also possible, such as failure to accurately treat the solvation free energy (which favors sheet); further simulations with water models other than TIP3P might help to address this possibility.

The effects of two simple and general perturbations to the potential, namely the removal of CMAP corrections and

extension of the short-range nonbonded cutoff, were considered in this study. While those perturbations failed to significantly alter the observed difference in free energies between SHEET and the helical states, they do illustrate how the results from DM can be used to test possible changes to the force field. Since the free energy difference between the native state and multiple decoy conformations can be calculated using the full potential in explicit solvent assuming some reference potential energy function, the free energy change for each of the reference states associated with perturbing the potential can then be used to determine whether that change favors the native state. If, as in the case of the WW domain, the native state is initially higher in free energy than some or all of the decoys, alterations to the potential to make it more stable than the other conformations would constitute a necessary (but not sufficient) modification to appropriately treat the folding and conformational equilibrium of the target protein. For example, approximations to the change in free energy in each reference conformation (see Corrective Perturbations to the Force Field in [Supporting Material](#)) show that altering the relative potential energy of the α - and β -region of the ϕ - ψ Ramachandran map by <0.45 kcal/mol would be enough to shift the balance between properly and improperly folded protein. Altering the hydration energy of amino acids by 13 kcal/mol would achieve a similar outcome, although in the latter case the magnitude of the needed perturbation appears to fall far beyond the acceptable range, based on previous studies indicating that current atomistic force fields (CHARMM, AMBER) are within 1–2 kcal/mol of experimental values (57,58). Additionally, such ad hoc corrections to the carefully parameterized CMAP and Lennard-Jones terms are unlikely to be generally applicable to systems other than the WW domain.

Improvement of a physics-based potential with implicit solvent to stabilize the native state of a protein has recently been used in tuning AMBER ff03 for use in structural refinement (17,59); similar tuning of the free energy differences between states using explicit solvent simulations based on deactivated morphing may also be useful for the refinement or testing of force fields for use in molecular-dynamics simulations. We must emphasize, however, that while the qualitative analysis of different components of the stability of the various conformations presented here can suggest factors that may be involved in the overstabilization of the helical states, it is not possible to unambiguously identify which of the competing physical effects involved is primarily responsible. It would, furthermore, be unrealistic to simply perturb single terms in the force field and attempt to identify one term to be corrected (for an example, see Corrective Perturbations to the Force Field in [Supporting Material](#)), because it is impossible to determine whether an altered term actually corrects an underlying physical defect in the force field or simply compensates for it sufficiently to stabilize the native state relative to misfolded states in the case of the WW domain. Such an effort would be

physically meaningful only in the context of testing a complete, systematically parameterized force field with changes expected to correct factors such as the hydrogen bonding geometry noted here, and would preferably be performed on a variety of different proteins. The lack of directionality in hydrogen bonding has also recently been suggested as a possible cause for the failure of modern MD force fields to yield appropriate thermodynamics for helix/coil equilibria in primarily helical peptides (R.B. Best and G. Hummer, unpublished).

The deactivated morphing procedure presented in Park et al. (27) and applied here provides a rigorous method for determining the free energy difference between two defined conformational ensembles of a biomolecule in explicit solvent. As detailed in The Deactivated Morphing Process in Supporting Material, ~300 ns of simulation were required for the transformation from $E(A)$ to $D(A)$ for each conformation A (see Methods), with an additional 50 ns of simulation for each morph between two conformations. While these computational costs are considerable, they are orders-of-magnitude less than the time that would be required to adequately sample both the helix and sheet structures in an equilibrium simulation (where SHEET was never observed in 21.9 μ s of simulation). Of the steps in DM, the restraining portion is by far the most expensive and carries the highest statistical uncertainty, so refinements to this portion of the procedure would be particularly useful. As in previous tests (27), it was found here that the relative free energy difference between two conformations due to the deactivation step can be approximated by the total internal potential energy of the protein, but unlike the case for decaalanine, the morphing component could not be adequately approximated by continuum electrostatics calculations for the WW domain.

In summary, to follow up on previous attempts to fold a fast-folding mutant of the human Pin1 WW domain in silico, we obtained three additional ~4 μ s MD trajectories of this protein, all of which formed similar distributions of misfolded helical states. Free energy calculations using deactivated morphing were then performed to determine the conformational free energy difference between the crystal structure and several commonly observed helical states, and indicated that the helical states are 4–8 kcal/mol more stable under these simulation conditions. Folding simulations of Fip35 will thus not be possible without alteration of the force field being used. As protein folding simulations and other long timescale MD simulations studying large conformational changes become increasingly common, force field parameterization or testing based on the establishment of proper free energy differences between chosen conformations of biomolecules become increasingly essential.

SUPPORTING MATERIAL

Four additional sections with figures and tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(09\)00588-8](http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)00588-8).

This work was supported by National Institutes of Health grant No. P41-RR05969 and National Science Foundation grant No. PHY0822613. Computer time was provided by the National Center for Supercomputing Applications through grant MCA93S028. P.L.F. has been supported by an National Science Foundation Graduate Research Fellowship.

The authors thank Dr. Chris Harrison for many useful discussions.

REFERENCES

- Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
- Lei, H., and Y. Duan. 2007. Ab initio folding of albumin binding domain from all-atom molecular dynamics simulation. *J. Phys. Chem. B.* 111:5458–5463.
- Duan, Y., and P. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1 microsecond simulation in aqueous solution. *Science.* 282:740–744.
- Simmerling, C., B. Strockbine, and A. E. Roitberg. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124:11258–11259.
- Chowdhury, S., M. C. Lee, G. Xiong, and Y. Duan. 2003. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J. Mol. Biol.* 327:711–717.
- Sanbonmatsu, K. Y., and A. E. Garcia. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins.* 46:225–234.
- Yang, W. Y., and M. Gruebele. 2004. Folding λ -repressor at its speed limit. *Biophys. J.* 87:596–608.
- Kubelka, J., T. K. Chiu, D. R. Davies, W. A. Eaton, and J. Hofrichter. 2006. Sub-microsecond protein folding. *J. Mol. Biol.* 359:546–553.
- Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding “speed limit”. *Curr. Opin. Struct. Biol.* 14:76–88.
- Maragakis, P., K. Lindorff-Larsen, M. P. Eastwood, R. O. Dror, J. L. Klepeis, et al. 2008. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J. Phys. Chem. B.* 112:6155–6158.
- Ensign, D. L., P. M. Kasson, and V. S. Pande. 2007. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J. Mol. Biol.* 374: 806–816.
- Freddolino, P. L., F. Liu, M. Gruebele, and K. Schulten. 2008. Ten-microsecond MD simulation of a fast-folding WW domain. *Biophys. J.* 94:L75–L77.
- MacKerell, A. D. 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584–1604.
- Wang, T., and R. Wade. 2006. Force field effects on a β -sheet protein domain structure in thermal unfolding simulations. *J. Chem. Theory Comput.* 2:140–148.
- Best, R. B., N.-V. Buchete, and G. Hummer. 2008. Are current molecular dynamics force fields too helical? *Biophys. J.* 95:L07–L09.
- Yoda, T., Y. Sugita, and Y. Okamoto. 2004. Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem. Phys.* 307:269–283.
- Wroblewska, L., and J. Skolnick. 2007. Can a physics-based, all-atom potential find a protein’s native structure among misfolded structures? I. Large scale AMBER benchmarking. *J. Comput. Chem.* 28:2059–2066.
- Zhou, R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins.* 53:148–161.
- Zhou, R., and B. J. Berne. 2002. Can a continuum solvent model reproduce the free energy landscape of a β -hairpin folding in water? *Proc. Natl. Acad. Sci. USA.* 99:12777–12782.
- Liu, F., and M. Gruebele. 2007. Tuning λ 6–85 towards downhill folding at its melting temperature. *J. Mol. Biol.* 370:574–584.

21. Nguyen, H., M. Jäger, J. W. Kelly, and M. Gruebele. 2005. Engineering a β -sheet protein toward the folding speed limit. *J. Phys. Chem. B.* 109:15182–15186.
22. Jäger, M., H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele. 2001. The folding mechanism of a β -sheet: the WW domain. *J. Mol. Biol.* 311:373–393.
23. Liu, F., D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, et al. 2008. An experimental survey of the transition between two-state and downhill protein folding scenarios. *Proc. Natl. Acad. Sci. USA.* 105:2369–2374.
24. Cecconi, F., C. Guardiani, and R. Livi. 2006. Testing simplified proteins models of the hPin1 WW domain. *Biophys. J.* 91:694–704.
25. Karanicolas, J., and C. L. Brooks. 2003. Improved Gō-like models demonstrate the robustness of protein folding mechanisms to non-native interactions. *J. Mol. Biol.* 334:309–325.
26. Luo, Z., J. Ding, and Y. Zhou. 2007. Temperature-dependent folding pathways of Pin1 WW domain: an all-atom molecular dynamics simulation of a Gō model. *Biophys. J.* 93:2152–2161.
27. Park, S., A. Y. Lau, and B. Roux. 2008. Computing conformational free energy by deactivated morphing. *J. Chem. Phys.* 129:134102.
28. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
29. Jäger, M., Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, et al. 2006. Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci. USA.* 103:10648–10653.
30. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.
31. Frishman, D., and P. Argos. 1995. Knowledge-based secondary structure assignment. *Proteins.* 23:566–579.
32. Roberts, E., J. Eargle, D. Wright, and Z. Luthey-Schulten. 2006. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics.* 7:382.
33. MacKerell, Jr., A. D., M. Feig, and C. L. Brooks, III. 2004. 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.* 25:1400–1415.
34. Andersen, H. C. 1983. RATTLE: a “velocity” version of the SHAKE algorithm for molecular dynamics calculations. *J. Chem. Phys.* 52:24–34.
35. Miyamoto, S., and P. A. Kollman. 1993. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins Struct. Funct. Gen.* 16:226–245.
36. van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, et al. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
37. Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, et al. 1999. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* 38:236–240.
38. Anitescu, M., and S. Park. 2009. A linear programming approach for the least-squares protein morphing problem. *Math. Program.* In press.
39. Singhal, N., C. D. Snow, and V. S. Pande. 2004. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper β -hairpin. *J. Chem. Phys.* 121:415–425.
40. Jayachandran, G., V. Vishal, and V. S. Pande. 2006. Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:164902.
41. Thompson, J. B., H. G. Hansma, P. K. Hansma, and K. W. Plaxco. 2002. The backbone conformational entropy of protein folding: experimental measures from atomic force microscopy. *J. Mol. Biol.* 322:645–652.
42. Lii, J.-H., and N. L. Allinger. 1998. Directional hydrogen bonding in the MM3 force field: II. *J. Comput. Chem.* 19:1001–1016.
43. Fabiola, F., R. Bertram, A. Korostelev, and M. S. Chapman. 2002. An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci.* 11:1415–1423.
44. Morozov, A. V., T. Kortemme, K. Tsemekhman, and D. Baker. 2004. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. USA.* 101:6946–6951.
45. Kortemme, T., A. V. Morozov, and D. Baker. 2003. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* 326:1239–1259.
46. Cramer, C., and D. Truhlar. 1999. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* 99:2161–2200.
47. Lum, K., D. Chandler, and J. Weeks. 1999. Hydrophobicity at small and large length scales. *J. Phys. Chem. B.* 103:4570–4577.
48. Huang, D. M., and D. Chandler. 2000. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. USA.* 97:8324–8327.
49. Wagoner, J. A., and N. A. Baker. 2006. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA.* 103:8331–8336.
50. Chen, J., and C. L. Brooks. 2008. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Phys. Chem. Chem. Phys.* 10:471–481.
51. Tanizaki, S., J. Clifford, B. D. Connelly, and M. Feig. 2008. Conformational sampling of peptides in cellular environments. *Biophys. J.* 94:747–759.
52. Feig, M. 2008. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* 4:1555–1564.
53. Harder, E., V. Anisimov, I. Vorobyov, P. Lopes, S. Noskov, et al. 2006. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* 2:1587–1597.
54. Weber, W., P. Hünenberger, and J. McCammon. 2000. Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B.* 104:3668–3675.
55. Hünenberger, P. H., and J. A. McCammon. 1999. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chem.* 78:69–88.
56. Kastenholz, M., and P. Hünenberger. 2004. Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods. *J. Phys. Chem. B.* 108:774–788.
57. Shirts, M. R., and V. S. Pande. 2005. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* 122:134508.
58. Deng, Y., and B. Roux. 2009. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* In press.
59. Jagielska, A., L. Wroblewska, and J. Skolnick. 2008. Protein model refinement using an optimized physics-based all-atom force field. *Proc. Natl. Acad. Sci. USA.* 105:8268–8273.