JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Sampling the Denatured State of Polypeptides in Water, Urea, and Guanidine Chloride to Strict Equilibrium Conditions with the Help of Massively Parallel Computers

Roberto Meloni,[†] Carlo Camilloni,[‡] and Guido Tiana*[,¶]

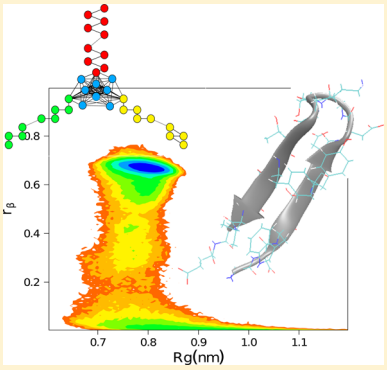
[†]Department of Physics, Università degli Studi di Milano, via Celoria 16, 20133 Milano, Italy

[‡]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

[¶]Department of Physics, Università degli Studi di Milano, and INFN, via Celoria 16, 20133 Milano, Italy

Supporting Information

ABSTRACT: The denatured state of polypeptides and proteins, stabilized by chemical denaturants like urea and guanidine chloride, displays residual secondary structure when studied by nuclear-magnetic-resonance spectroscopy. However, these experimental techniques are weakly sensitive, and thus molecular-dynamics simulations can be useful to complement the experimental findings. To sample the denatured state, we made use of massively-parallel computers and of a variant of the replica exchange algorithm, in which the different branches, connected with unbiased replicas, favor the formation and disruption of local secondary structure. The algorithm is applied to the second hairpin of GB1 in water, in urea, and in guanidine chloride. We show with the help of different criteria that the simulations converge to equilibrium. It results that urea and guanidine chloride, besides inducing some polyproline-II structure, have different effect on the hairpin. Urea disrupts completely the native region and stabilizes a state which resembles a random coil, while guanidine chloride has a milder effect.



INTRODUCTION

A large number of experiments agree on the fact that the denatured state of proteins and peptides is usually not a random coil but contains residual structure.[1,2] Being marginally stable, the experimental characterization at the atomic level of such a residual structure is a formidable task. In particular, the only approaches that can give detailed information of the location of structured regions in the denatured state are those based on nuclear magnetic resonance,[3−5] which is a weakly-sensitive technique. As a consequence, our understanding of the denatured state of proteins is very limited with respect to that of the folded state.

In this respect, molecular-dynamics (MD) simulations are expected to be a useful complement to experimental data, because they can provide equilibrium averages of any quantity of interest both in water and in a large variety of denaturants, including the widely-used urea and guanidine chloride. The limit of plain MD simulations is that the computational time needed to achieve equilibration can be exceedingly large even for top supercomputers,[6−8] because biomolecules usually display relevant thermodynamic states separated by large (i.e., > kT) free-energy barriers. As a consequence, smarter algorithms were developed, like Umbrella Sampling (US),[9] Replica-Exchange MD (REX),[10] Hamiltonian REX,[17] and Metadynamics[11] to fasten the crossing of such barriers. Moreover, these techniques were used in combination with each other, like for example combining REX and metadynamics[12] or REX and US.[21,22]

The goal of the present work is to sample the denatured state of a polypeptide, comparing its properties in water, in urea, and in guanidine chloride. In the literature, the denatured state often is described as an intrinsic property of the polypeptide, with little emphasis on the denaturant which stabilizes it. Through a comparison between the structural features of the denatured state in two different denaturing agents, and between them and the metastable denatured state in water, we intend to investigate the role of the denaturant in defining the denatured state. For this purpose, we make use of a variant of the Hamiltonian REX algorithm, biased toward the formation and disruption of secondary structure elements, and thus particularly suitable for this specific problem. In fact, the denatured state of polypeptides is characterized by a multiplicity of conformations separated by small free-energy barriers and by residual secondary structure, as highlighted by nuclear-magnetic-resonance experiments in denaturing conditions.[2,13−16] As a consequence, the associated computational problem is not much the crossing of few large free-energy barriers but that of enhancing molecular diffusion to sample a large conformational space, visiting a large amount of conformations characterized by different secondary structures.

In standard Hamiltonian REX,[17] one simulates M replicas of the system, each coupled with a heat bath at temperature T_i and interacting through its own Hamiltonian H_i. Each replica can exchange its conformation with the other replicas with a rate

Received: October 8, 2013
Published: December 20, 2013

chosen to keep the system in a multicanonical ensemble. Usually, the potential in the Hamiltonians H_i is in the form $U_i = U_0 + \lambda_i V$, where U_0 is the standard molecular potential, V is a biasing potential, and λ_i is the relative weight of the bias. The biasing potential can be used to force the different replicas toward or against specific states, for example to favor specific values of Ramachandran dihedrals,[18] to foster the formation of specific contacts,[19] or to weaken interactions that slow down conformational sampling.[20] Exchanges are usually attempted between neighboring replicas, in order to maximize the exchange rate, producing a linear arrangement of the connection between replicas. In our implementation a core of unbiased replicas is connected to three branches of replicas (Figure 1), biased to favor the formation of alpha-helices, of

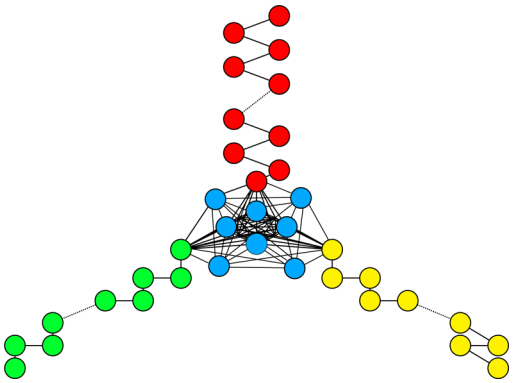


Figure 1. The topology of the replica network used for Hamiltonian replica exchange simulations. Blue circles represent unbiased replicas at 300 K, from which the α, the β, and the temperature replicas branch off.

beta-structures, and the disruption of formed structures, respectively. The unbiased replicas, exchanging their conformation with the three branches, are expected to sample efficiently their conformational space. Moreover, reweighting the conformations sampled in the three branches, in order to remove the effect of the bias, further improves the sampling statistics. This approach is particularly effective if used on massively parallel computers like Blue-Genes, where it is possible, and often mandatory, to use thousands of processors. Indeed, all REX algorithms scale linearly with the number of processors, the computational overhead due to the exchange between replicas being usually negligible, and thus are particularly suitable for massively-parallel computers. A nice feature of our algorithm is that, similarly to Bias-Exchange Metadynamics,[23] it scales linearly also with respect to the number of biasing variables, whereas REX/US scales exponentially with it.[22]

In the following, we shall study the conformational space of the second hairpin of protein GB1, a small system which has been extensively characterized both experimentally[24−26] and computationally.[27−30] The study is first carried out under native conditions, where the hairpin is stable, and we can compare the results of our calculations with those reported in the literature. We provide evidence that the present REX scheme can reconstruct efficiently the equilibrium population of this polypeptide, exploiting the features of massively-parallel computers. A detailed statistical analysis of the convergence of the simulation to equilibrium conditions is carried out with several methods, including a bootstrap approach. After comparing the behavior of the peptide in water with that described in the literature, we focus our attention to the case of two denaturing solutions, obtained adding to the system urea and guanidine hydrochloride (GndCl), respectively. The goal is two-fold. First, this is the first step in the study of the denatured state of larger polypeptides. Moreover, it is a simple system in which one can study the denaturing effect of two popular chaotropic agents under equilibrium conditions.

METHODS

The second hairpin of GB1, extracted from the pdb structure 1PGB leaving the termini uncapped, is solvated in a cubic box of side 4.82 nm with 3620 TIP3p water molecules and 3 Na^+ ions. The atoms interact through the Amber99 potential,[31] modified as in ref 32 in order to have a better balance between the secondary-structure propensities of the chain. The simulations are carried out with Gromacs 4.5.5[33] coupled to a modified version of Plumed 1.3[34] at fixed volume, with Nosé−Hoover thermostat and Particle Mesh Ewald treatment of the electrostatic interactions. The simulations carried out in denaturing conditions make use either of the standard Amber urea parameters or of the parameters for guanidine hydrochloride described in ref 35. In the case of urea, the hairpin is solvated in a cubic box of side 4.68 nm with 2524 TIP3p water molecules, 3 Na^+ ions, and 338 urea molecules, corresponding to a 5.5 M solution; in guanidine chloride solution, the cubix box side is 4.78 nm, and there are 2705 TIP3p water molecules, 203 guanidine molecules, corresponding to a 3.8 M solution, 3 Na^+ ions and 203 Cl^- ions to ensure charge neutrality.

Each simulation comprises 128 replicas of the system, running on a total of 1024 cores. Of them, $n_0 = 8$ are unbiased (replicas 0), $n_\alpha = 20$ are biased toward the formation of a α-helix (replicas α), $n_\beta = 50$ are biased toward the formation of β-hairpins (replicas β), and $n_T = 50$ are unbiased but are simulated at increasing temperatures (replicas T). Replicas 0, α, and β are simulated at temperature $T = 300$ K. In replicas α and β, each replica i the potential has the form $U_i = U_0 + \lambda_i V$, where U_0 is Amber potential and

$$V(r) = \frac{1}{2}(r-1)^2 \tag{1}$$

where r denotes a collective coordinate of the system, which is the relative degree of helicity for replicas α (denoted as r_α in the following) described in refs 34 and 36 and normalized by 10, in order to have $r_\alpha = 1$ for a perfect helix, and the relative degree of beta structures for replicas β (denoted as r_β in the following) described in refs 34 and 36 and normalized by 4, in order to have $r_\beta = 1$ for a perfect beta-hairpin.

Pairs of replicas i and j which are connected according to the scheme displayed in Figure 1 can exchange their conformation X and X' with rate

$$w(X, i; X', j) \propto \min\left(1, \exp\left[-\frac{U_j(X')}{kT_j} - \frac{U_i(X)}{kT_i} + \frac{U_j(X)}{kT_j} + \frac{U_i(X')}{kT_i}\right]\right) \tag{2}$$

The values of λ_i range from 0.3 to 18 in replicas α and from 1 to 50 in replicas β, while the temperatures range from $T = 300$ K to $T = 484$ K in replicas T; the choice of temperatures was performed in such a way to ensure uniform exchange probabilities.[37]

Each replica is first thermalized with 1 ns of simulation without allowing exchange between replicas. The production simulations are run on a Blue Gene/Q machine, with each

replica parallelized over 8 cores. This results in a computational cost of approximately 4 ns/day per replica.

Free energies are reconstructed with a weighted-histogram algorithm.[38−40] To match the information provided from the biased replica (called $\mathcal{B}$) and the temperature replica (called $\mathcal{T}$), we start from the standard weighted-histogram equations[38] for the unbiased probability $p_0(r)$ of the generic (usually multidimensional) variable r at inverse temperature β_0

$$p_0(r) = \frac{\sum_i n_i(r)}{\sum_i N_i Z_i w_i(r)} \tag{3}$$

$$Z_i^{-1} = \sum_r w_i(r) p_0(r) \tag{4}$$

where $n_i(r)$ is the histogram as calculated by the ith replica of the simulation, $N_i = \sum_r n_i(r)$, $w_i(r)$ are the reweighting factors

$$w_i(r) = \begin{cases} \exp[-(\beta_i - \beta_0)U(r)] & \text{if } i \in T \\ \exp[-\beta_0 \lambda_i V(r)] & \text{if } i \in B \end{cases} \tag{5}$$

Calling $U_i^m(r)$ the minimum energy visited by the ith replica and $Z_i' \equiv Z_i e^{-(\beta_i - \beta_0)U_i^m}$ if $i \in \mathcal{T}$ and $Z_i' \equiv Z_i$ if $i \in \mathcal{B}$, the above equations are rewritten as

$$p_0(r) = \frac{\sum_i n_i(r)}{\sum_{i\in B} N_i Z_i' \exp[-\beta_0 \lambda_i V(r)] + \sum_{i\in T} N_i Z_i' \exp[-(\beta_i - \beta_0)(U(r) - U_i^m)]}$$

$$(Z_i')^{-1} = \begin{cases} \sum_r \exp[-(\beta_i - \beta_0)(U(r) - U_i^m)] p_0(r) & \text{if } i \in \mathcal{T} \\ \sum_r \exp[-\beta_0 \lambda_i V(r)] p_0(r) & \text{if } i \in \mathcal{B} \end{cases} \tag{6}$$

that can be solved recursively to find Z_i'. The change of variables from Z_i to Z_i' prevents numerical overflows in the calculation, because it constrains the arguments of the exponentials to negative values. To solve eqs 6, we have used a modified version of the software described in ref 41. Thermal averages of variables depending on the conformation of the chain are calculated integrating the variable, weighted by the probability found in eq 6.

Clustering is carried out first merging together all conformations sampled in all the replicas, clustering this set with a linkage algorithm on the mutual RMSD with a cutoff at which the number of clusters is half of the number of sampled conformations, and reweighting each cluster according to the sum of the weigths of the associated conformations.

RESULTS

Study of the Convergence Properties of the Simulations. The convergence of the simulation was assessed by several methods. We started from the hairpin in water, in which it is expected to populate largely its native conformation.[24] We first inspected the one-dimensional free energies (see Figure S1 in the Supporting Information), calculated at 300 K with the weighted-histogram method making use of the conformations generated from the simulation up to different times and projected on r_α, r_β, and the gyration radius R_g, respectively. The free energy profile does not change any longer after 45 ns, something which is a necessary condition for convergence.

A standard way of assessing the convergence of a free-energy is to compare the profiles obtained from the second third and from the last third of the simulation. In Figure 2 it is plotted the free energy calculated with the weighted-histogram method making use of the conformations sampled in the second third (from 20 to 40 ns) and in the last third (from 40 to 60 ns) of the simulation, for r_α, r_β, and R_g. The first third is discarded to prevent dependence on the initial conditions. All the pairs of profiles display almost everywhere a difference within ≈1 kJ/mol ≈ 0.4 kT, with the exceptions of the very-high-energy regions which are not visited exhaustively by any replica. More precisely (see Figure S2 in the Supporting Information), a difference of more than 1 kJ/mol between the two free energy profiles affects the points whose free energy is above 12.5 kJ/mol from the minimum in the case of r_α, none of them in the case of r_β, and the points above 22 kJ/mol from the minimum in the case of R_g.

Another tool to estimate the convergence of a simulation is to compare the cluster of conformations sampled in the second and in the last third of the simulation.[42] Making use of the scheme discussed in the Methods section, we clusterized the conformations and associated with them a statistical weight. The difference between the statistical weights of the corresponding clusters is 0.08 for the "native" cluster (i.e., the associated statistical weight is 0.89 in the second third and 0.81 in the last third of the simulation) and smaller than 0.01 for the other clusters (see also Figure S3 in the Supporting Information).

To get a deeper understanding of the degree of convergence of the simulations, we also investigated a further criterion, based on a bootstrap scheme. We first divided the replicas into two interwined groups, called "even" and "odd" in the following. From each of the two groups we extract a set of uncorrelated conformations, apply the weighted histogram, and project them to obtain the free energy profile as a function of r_α, of r_β, and of R_g. The difference between the two free-energy profiles is then compared with 200 profiles obtained resampling at random the conformations of each replica, applying to them the multiple histogram algorithm, thus reconstructing the one-dimensional free energy as a function of r_α, r_β, or R_g. The difference between even and odd free energies are within the standard deviation of the bootstrapped free energies (see Figure S4 in the Supporting Information).

By splitting the sampled conformations into two uncorrelated sets according to the replicas they come from (even and odd) instead of according to the time block, we aim at making the most of the available information in the two sets. This is important in the present case, in which sampling is mostly diffusion limited, and it is possible due to the large amount of

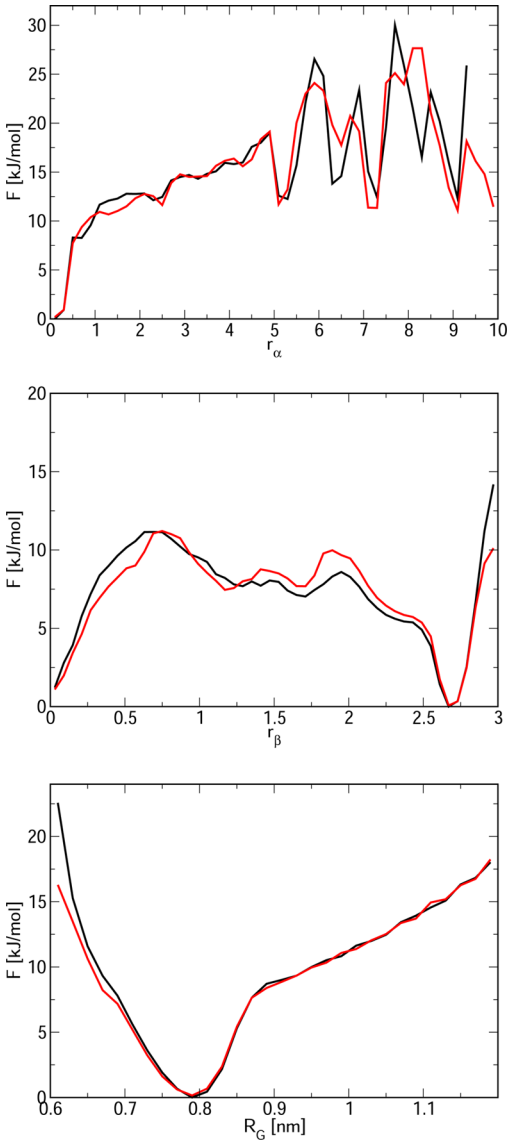


Figure 2. The free energy calculated with the weighted-histogram method as a function of r_α (upper panel), r_β (middle panel), and of the gyration radius R_g (lower panel) from the conformations sampled in the second third (black curve) and in the last third (red curve) of the simulation.

replicas involved in the calculation. In fact, when comparing the free energies obtained by time blocks (as done above), one risks not giving the system enough time to diffuse in conformational space.

The requirement that the two sets are uncorrelated is critical, to be sure that the two distributions are not trivially similar. For this purpose, we study the cross-correlation

$$C_{i,j}(\tau) = \langle r_i(t) r_j(t+\tau)\rangle_t - \langle r_i(t)\rangle_t - \langle r_j(t)\rangle_t \quad (7)$$

where the variable r indicates r_α, r_β, or R_g calculated along fixed adjacent replicas i and j which can exchange conformation with each other, and $\langle\cdot\rangle_t$ indicate average over the simulation time (cf. Figure S5 in the Supporting Information). For the analysis, we retain conformations for which the normalized cross-correlation is smaller than $e^{-1} \approx 0.3$. The cross-correlation time τ_{ij}, which reflects the diffusion of conformations among different replicas, is typically some tens of picoseconds, except for the temperature branch, where it can reach the nanoseconds scale. Moreover, to implement the bootstrap method on the correct number of uncorrelated conformations, we first calculate the correlation time τ_i of each replica, defined as the time at which both the auto-correlation functions of r_α and r_β drop below e^{-1}, thanks mainly to the conformational exchange with other replicas. Then we prune the trajectory in each replica, retaining one conformation every τ_i and making sure that in each replica there are at least 100 conformations; see Figure S6 in the Supporting Information.

To be more quantitative in the bootstrapping algorithm, we reconstruct the probability distributions $p_{even}(r)$ and $p_{odd}(r)$ from the even and from the odd free energies, we compare them with the Jensen–Shannon distance,[43] defined as

$$\begin{aligned} d_{JS}[p_{even}(r), p_{odd}(r)] = {} & \frac{1}{2}\int_{-\infty}^{+\infty} dr[p_{even}(r)\log(p_{even}(r))] \\ & + \frac{1}{2}\int_{-\infty}^{+\infty} dr[p_{odd}(r)\log(p_{odd}(r))] \\ & - \frac{1}{2}\int_{-\infty}^{+\infty} dr(p_{even}(r) \\ & + p_{odd}(r))\log\left[\frac{1}{2}(p_{even}(r) + p_{odd}(r))\right] \end{aligned} \quad (8)$$

and compare the resulting distance with those obtained by bootstrapping. This procedure is applied separately to r_α, r_β, and R_g, and r in eq 8 indicates generically these variables. As we show in the following, this is a very sensitive criterion, which is useful in our specific problem because our goal is to investigate residual secondary structure which is only marginally populated in the denatured state, and consequently we need a high degree of convergence.

The result is $d_{JS} = 0.0064$ in the case of r_α, $d_{JS} = 0.0009$ in the case of r_β, and $d_{JS} = 0.0045$ in the case of R_g. This result is compared with the result obtained by bootstrapping in Figure 3

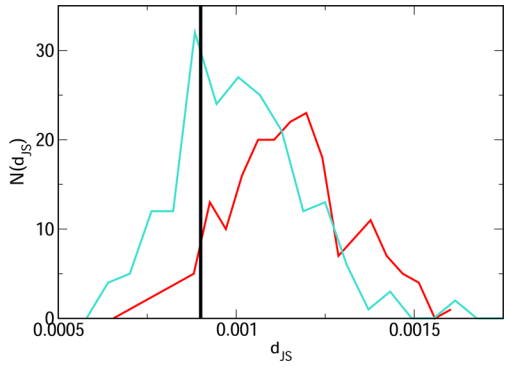


Figure 3. The Jensen–Shannon distance between the distributions $p_{even}(r_\beta)$ and $p_{odd}(r_\beta)$ (marked by a black line) and the histograms N of the Jensen–Shannon distances calculated between $p_{even}(r_\beta)$ and the distribution $p_{boot-even}(r_\beta)$ obtained with 500 bootstraps (red curve) and those calculated between $p_{odd}(r_\beta)$ and $p_{odd-even}(r_\beta)$ (cyan curve).

for the case of the probability as a function of r_β (and in Figure S7 of the Supporting Information for r_α and R_g). The Jensen–Shannon distance between the even- and odd-replica distributions lies at the center of the distribution of bootstrapped distances, indicating that even- and odd-replica are not more different than the precision with which we can obtain the distribution probability with the weighted-histogram method.

dx.doi.org/10.1021/ct400879v | J. Chem. Theory Comput. 2014, 10, 846–854

Specifically, it is within $0.52\sigma_e$ and $0.68\sigma_o$ from the average of the bootstrap distribution, where σ_e and σ_o are the standard deviation of the bootstrapped distribution for the even and odd replicas, respectively.

Interestingly, this way of assessing the convergence is more sensitive than the standard comparison of free energies carried out above. In fact, keeping only the first 40 ns of the simulation and carrying out the standard analysis on the free energies, one obtains a result which is very similar to that obtained making use of the whole simulation (cf. Figure S8 in the Supporting Information). On the other hand, carrying out the Jensen−Shannon analysis as described above (cf. Figure S9 in the Supporting Information) highlights a serious lack of convergence.

Similar simulations have been carried out in 5.5 M urea and in 3.8 M GndCl. The time needed to reach equilibration is somewhat smaller than that in water (cf. Figures S10 and S11 in the Supporting Information). The reason is likely to be found in the lack of a two-state behavior and thus in the absence of a major barrier between the two states. The comparison between even and odd replicas is shown in Figure 4 for the two

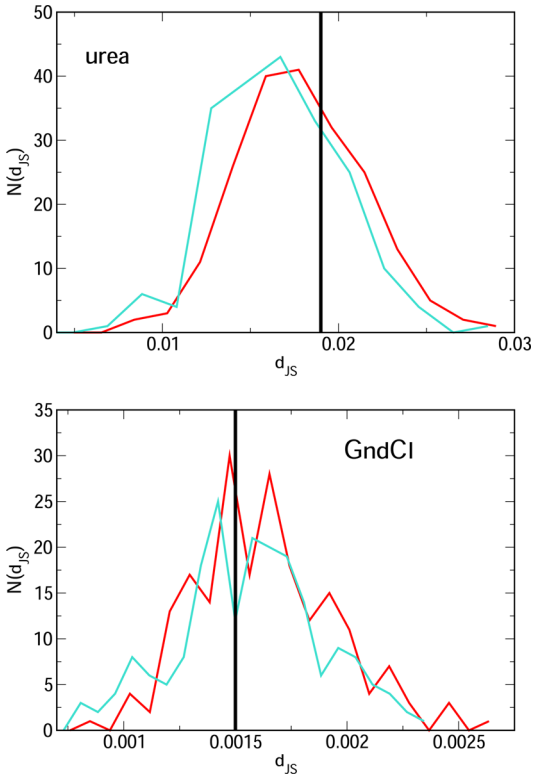


Figure 4. The Jensen−Shannon distance for the simulations in urea and GndCl, as calculated in Figure 3.

denaturants, making again use of the Jensen−Shannon distance between the reconstructed distributions of r_β (the comparison between the distributions of r_α and R_g are displayed in Figures S12 and S13 in the Supporting Information). The differences between the two distributions are within $0.31\sigma_e$ and $0.37\sigma_o$ in the case of urea and $0.08\sigma_e$ and $0.11\sigma_o$.

Properties of the Hairpin in Water. The free energy of the hairpin in water as a function of r_β and R_g at 300 K is displayed in Figure 5. The system displays the native hairpin as global minimum, a denatured state lacking of beta structure

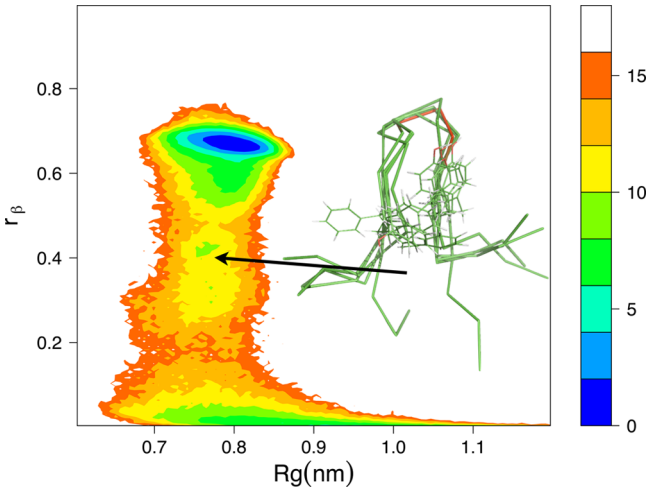


Figure 5. The free energy (in kJ/mol) as a function of the gyration radius R_g and of the degree of beta structure r_β for the hairpin in water at 300 K. In the inset, five representative conformations of the intermediate state; TYR45 and PHE52 are highlighted in the structure.

and spanning a wide range of gyration radii, and a high-energy intermediate state. The intermediate state, which lies ≈ 10 kT above the native state, is composed of conformations which are hairpin-like in the central region, between TYR45 and PHE52 (which are highlighted in the figure), while the termini are disordered.

The thermal average of r_β as a function of the temperature is displayed in Figure 6 as a black curve. At 300 K the average

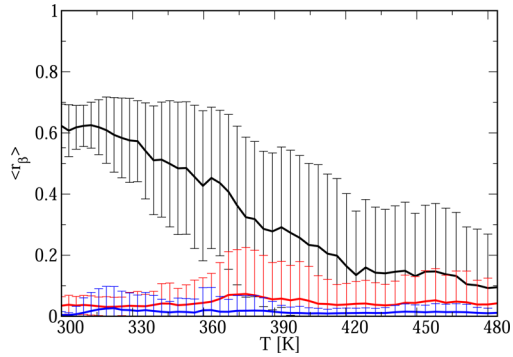


Figure 6. The average degree of β-structure r_β as a function of temperature calculated from the simulation in water (black curve), urea (blue curve), and GdnCl (red curve). The error bars indicate the standard deviation.

degree of formation is ≈0.6, comparable with the experimental value 0.5 obtained by fluorescence measurements.[25] The curve decreases quite slowly in temperature and retains a sizable amount of beta structure even at large temperatures. The average gyration radius, reported in Figure 7, shows a limited variation as a function of temperature, remaining close to its native value of 0.81 nm up to ≈400 K, and then having a small jump to 0.85 nm. The amount of α-helix is null at 300 K and remains within 10% at larger temperatures (cf. Figure S14 in the Supporting Information).

The denatured state, operatively defined as the set of conformations with $r_\beta < 0.25$ (cf. Figure 5), has been further analyzed. Its gyration radius is 0.83 ± 0.07 nm, very similar to the actual thermal average, calculated over the whole

dx.doi.org/10.1021/ct400879v | J. Chem. Theory Comput. 2014, 10, 846−854

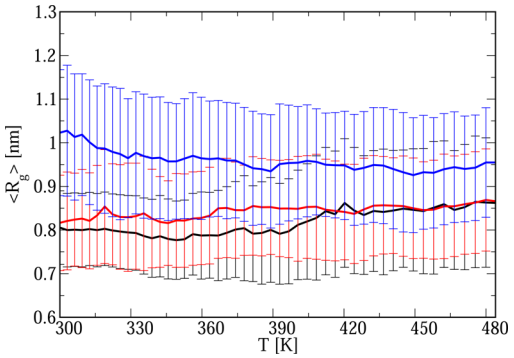


Figure 7. The average gyration radius R_g as a function of temperature calculated from the simulation in water (black curve), urea (blue curve), and GdnCl (red curve). The error bars indicate the standard deviation.

conformational space. The distribution of root-mean-square difference (RMSD) between each pair of denatured conformation, calculated reweighting all the conformations of all replica, is displayed with a black curve in Figure 8. The

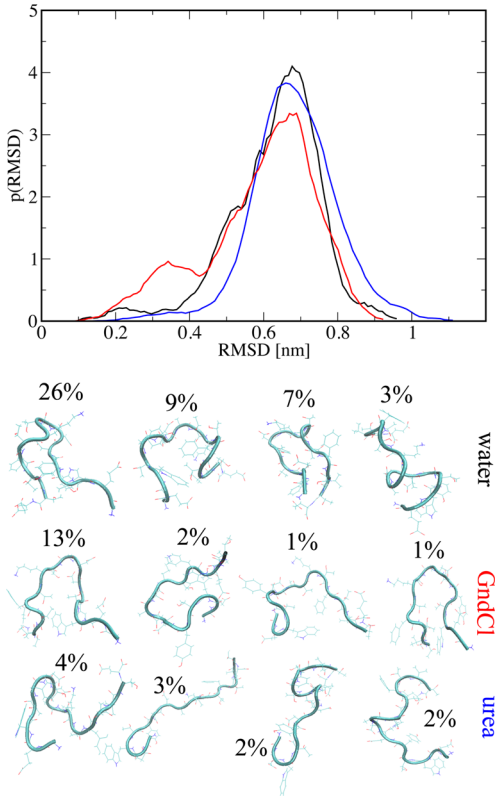


Figure 8. In the above panel, the distribution of RMSD between the pairs of conformations at 300 K which populate the denatured state (r_β < 0.25) in water (black curve), urea (blue curve), and GdnCl (red curve). The conformations to the right are the representatives of a clustering analysis of the associated conformations.

distribution displays a main peak centered at an RMSD of ≈0.7 nm, corresponding to pairs of conformations completely different from each other, and a tail toward small values of RMSD. This tail indicates that there are clusters of conformation that, at equilibrium, are populated by structurally-similar conformations. A clustering analysis has being carried out to identify them, assigning two conformations to the same cluster if their RMSD is lower than a threshold. The threshold is chosen is such a way that the number of clusters is half of the number of conformations (see Figure S15 in the Supporting Information). The representative conformations of the more populated clusters are displayed in the lower panel of Figure 8. The system populates a main cluster displaying a probability of 26% in which the N-terminus of the peptide forms a U-like conformation. Two other clusters, both with some sort of U-like shape, are populated at 9% and 7%, while the other clusters have probabilities of at most a few percents.

Secondary structure populations, assigned using a modified version of STRIDE that takes into account PolyProline II (PPII) geometries,[15,44,45] for residues in the conformations with r_β < 0.25 are calculated integrating the associated probabilities, calculated with the WHAM algorithm, and are displayed in Figure 9. It results that the only turns are present in the denatured state.

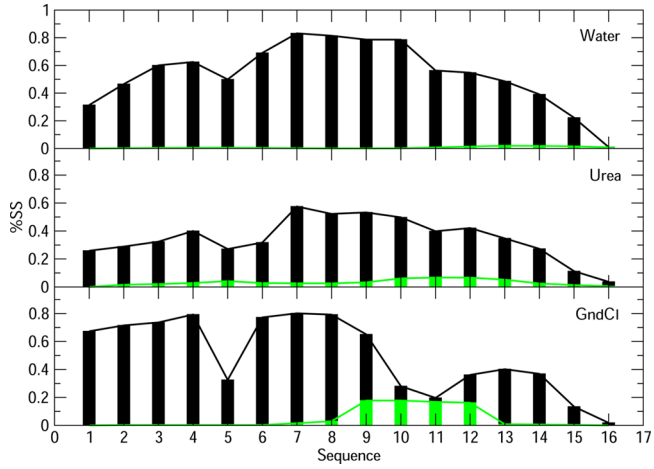


Figure 9. The amount of secondary structure per residue (turn indicated by black bars, while polyproline-II (PPII) indicated by green bars), calculated in water (top), urea (middle), and GdnCl (bottom panel).

Properties of the Hairpin in Urea and Guanidine Chloride. The properties of the peptide in water are then compared with those in urea and GdnCl solutions. The free energy as a function of R_g and r_β is displayed in Figure 10 for the simulations of the peptide at 300 K in the two denaturants. In the case of the urea solution, the free energy does not display a well-defined mimimum in the native region, but a ragged profile whose free energy is ≈10 kT larger than that of the completely-denatured state. As a consequence the degree of formation of the β-hairpin is nearly zero at 300 K and remains low at all temperatures, as shown with a blue curve in Figure 6. The gyration radius of the peptide in urea at 300 K is much larger than that in water, being slightly larger than 1 nm, and decreases to 0.95 nm when the temperature is raised (cf. blue curve in Figure 7). The amount of α-helix as a function of temperature (cf. blue curve in Figure S14 in the Supporting Information) is very similar to that of the peptide in water, slightly increasing with temperature, but staying below 10%.

The effect of GdnCl on the peptide is markedly different. Its free energy, displayed in the lower panel of Figure 10, has the same minima as that in water. The native minimum has a higher free energy and the intermediate a lower free energy than that in water, but both of them are still well-defined. As a

dx.doi.org/10.1021/ct400879v | J. Chem. Theory Comput. 2014, 10, 846−854

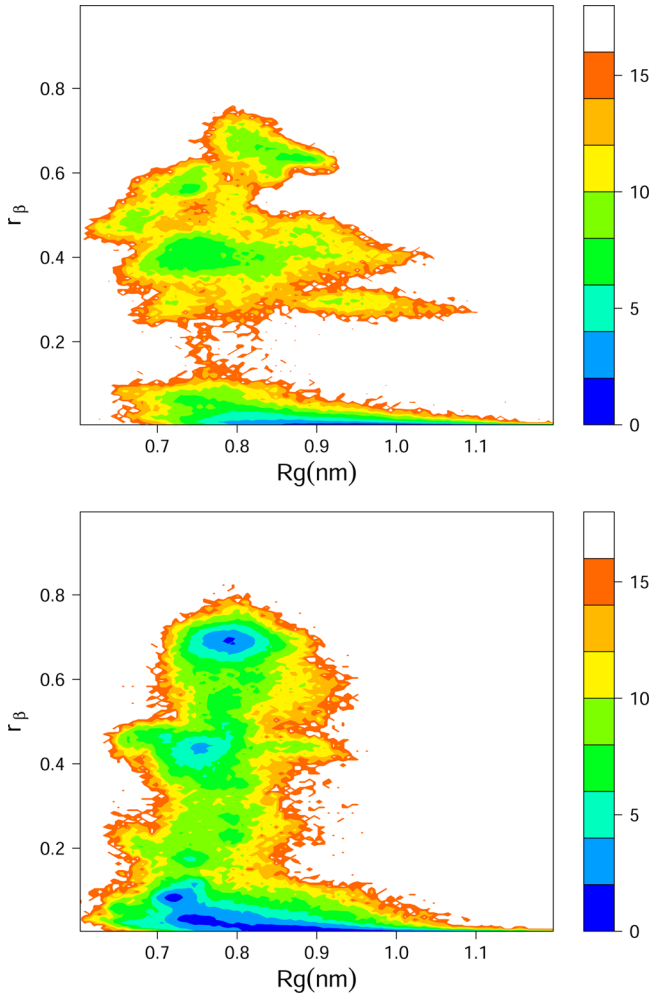


Figure 10. The free energy (in kJ/mol) as a function of the gyration radius R_g and of the degree of beta structure r_β for the hairpin in urea (above) and GndCl (below) at 300 K.

result, the average degree of β-structure is around 5% and remains pretty constant at all temperatures (see the red curve in Figure 6). The gyration radius, displayed as a red curve in Figure 7, is slightly larger, but comparable, to that in water at 300 K, and converged to it as temperature is raised. The amount of α-helix increases with temperature from ≈5% to ≈10% (cf. Figure S14 in the Supporting Information).

The denatured state (r_β < 0.25) is then analyzed and compared with that of the system in water. Both the distributions of RMSD in urea and GndCl display a peak around 0.7 nm, indicating that most of the sampled conformations are very different from each other. However, the peak in GndCl is more broadened toward small values of RMSD and also displays a further peak around 0.3 nm, similarly but to a larger extent than in water, indicating the presence of a set of conformations that at equilibrium are self-similar. These data suggest that the denatured state in GndCl and in water are more structured than in urea.

Repeating the cluster analysis in the simulation under denaturing conditions, it is found that in GndCl the system populates a cluster of U-like conformations with probability 13%, and all the others with probability lower than a few percents. In urea there is not any dominant cluster, all of them accounting for a few percents of the whole probability.

This difference is also apparent in the distribution of gyration radii (Figure S11 in the Supporting Information). While in urea the distribution of gyration radius is broadened between 0.8 and 1.3 nm, in GndCl it displays a sharp peak at 0.75 nm, corresponding to the most populated cluster of conformations discussed above, and a tail at large values. Interestingly, this difference is not reflected by the end-to-end distance, that displays in both cases a broad distribution between 0.2 and 3 nm (see dashed curve in Figure S16). Consequently, the end-to-end distance of the hairpin, frequently mapped in fluorescence-transfer experiments,[25] does not seem a good order parameter to study the states of this system.

The peptide is rich of β-turn in all solvents; the fractional amount of β-turn is larger toward the center of the peptide and decreases toward the termini, with a drop corresponding to residue Y5. The maximum reached in water and in GndCl is similar to each other (≈0.8) and is slightly larger than in urea (≈0.6). However, the behavior in urea displays a marked difference with respect to the other solvents in the region 9−12, in which the amount of β-turn drops and is substituted by PPII with a probability ≈0.2. While in GndCl only this region displays PPII structure, in urea this is more diffuse, involving, to a minor extent (≈0.1), the whole chain. In water the amount of PPII is negligible. These results are in agreement with the observation by circular dichroism that both urea and GndCl promote the formation of PPII.[46] The probability of the other kinds of secondary structure is essentially zero and consequently is not shown in the figure.

DISCUSSION

The availability of parallel computers with a large number of computing nodes ($\gtrsim 10^3$) cannot be a real breakthrough in the study of the conformational space of macromolecules if not accompanied by the development of suitable algorithms. To take full advantage of the power of massively parallel computers we suggest a strategy based on a multibranch replica exchange algorithm, in which the different branches, connected with some unbiased replicas, are meant to favor the formation and disruption of local secondary structure. In this way we are able to sample efficiently the denatured states of polypeptides, which are usually characterized by transient secondary structures. The equilibrium properties of such systems is recovered from the analysis of the unbiased replicas, combined through a weighted-histogram calculation with that of the biased replica branches. As a result, we can reconstruct the free-energy profile of a 16-residue peptide in water in ≈60 ns per replica using 128 replicas on 1024 cores, to be compared with 500 ns on 32 replicas needed for a REX simulation in water[30] and 100 ns on 64 replicas for a REX-metadynamics simulation in urea.[47] Although we are not able to assess the degree of convergence of these samplings, their computational cost seems comparable to ours. However, our variant of the Hamiltonian REX algorithms seems particularly well suited for machines with a large number of relatively slow processors. Since the availability and the cost of computers with many nodes is usually more accessible than that of machines with a smaller number of ultrafast processors, we believe that the development of algorithms for the formers can be useful to study the equilibrium properties of biological macromolecules. Moreover, to expand this kind of calculation to larger systems, it is certainly easier to increase the number of nodes than the speed of the processors. Furthermore when using large-computing facilities it is important to achieve as far as possible a linear scaling of the sampling performances with the number of processors. Indeed, replica-exchange is a linear scaling algorithm while the scaling of algorithms used to parallelize a single

simulation, like domain decomposition, are always limited by the density of the system and the range of the interactions.

A common problem when studying the equilibrium properties of a system is to assess to what extent the sampling is converged.[6] For this purpose, we have also developed a statistically-sound but economic strategy to evaluate if the quantities of interest are equilibrated. Based on a bootstrap method, it quantifies to which extent the distributions of the variables under study, calculated with the weighted-histogram method applied to uncorrelated sampled conformations are similar, as compared to the reproducibility of the weighted-histogram method itself. With this algorithm we have shown that we have reached equilibrium of the quantities of interest (i.e., degree of secondary structures and gyration radius) within some fraction of the error with which the WHAM can determine them.

The peptide we used to test the algorithm was chosen because it is widely studied in the literature. In particular, it was shown in ref 30 that its equilibrium properties, as calculated by MD simulations, are strongly dependent on the force field used in the simulation. For this reason we used the Amber99SB*-ILDN-Q force field,[32] which represents one of the latest developments of force-field tailored to correctly represent the secondary structure content of biomolecules.[48,49] Interestingly, making use of the Amber99SB*-ILDN-Q force field we obtain in water an intermediate state similarly to what was observed with OPLS[29] and Amber99SB; however, the structural properties of the intermediate are different. In OPLS it is a non-native out-of-register hairpin, in AMBER99SB it is native-like toward the termini, while we find with Amber99SB*-ILDN-Q a set of conformations with the native hairpin formed in the half close to the turn. On the other hand, the strong effect of urea on the structure and on the compactness of the peptide using the Amber99SB*-ILDN-Q force field seems comparable to what was found in OPLS,[47] although in the latter study a higher concentration of urea was employed.

The main result of the simulations is that the peptide in GndCl has properties which are much more similar to the denatured state in water than in urea. For the specific case of this β hairpin, urea seems to display a stronger denaturing effect than GndCl. In urea, the native population of the peptide is smaller, and the denatured state is more disordered and less compact, behaving similarly to a random coil. On the contrary, in GndCl the population of the native state is larger, and also the denatured state is far from a random coil. It is well compact and populate consistently loop-like conformations. This difference contrasts with the popular idea that GndCl is overall a stronger denaturant than urea and points to more specific differences. The destabilizing effect of urea on hairpins is in agreement with simulations of denaturation of protein L,[35] whose sequence is unrelated to that of protein G, carried out with a completely different force field, that is Gromos96. In that case, it was observed that urea first disrupts the β-hairpin of protein L, while GndCl first disrupts the α-helix. On the other hand, both urea and GndCl induce the formation of PPII structure, as already observed in ref 46. This effect is more sequence-specific in the case of GndCl and more diffuse along the sequence in the case of urea.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S16. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tiana@mi.infn.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the computational support of CINECA concerning the use of the Blue Gene machine "Fermi". C.C. was supported by a Marie Curie Intra European Fellowship.

■ REFERENCES

(1) Dill, K. A.; Shortle, D. Denatured states of proteins. *Annu. Rev. Biochem.* **1991**, *60*, 795.

(2) Bowler, B. E. Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol.* **2012**, *22*, 4.

(3) Dyson, H. J.; Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **2004**, *104*, 3607.

(4) Vendruscolo, M. Determination of conformationally heterogeneous states of proteins. *Curr. Opin. Struct. Biol.* **2007**, *17*, 15.

(5) Mittag, T.; Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.

(6) Zuckerman, D. M. Equilibrium sampling in biomolecular simulations. *Ann. Rev. Biophys.* **2011**, *40*, 41.

(7) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **2012**, *134*, 3787.

(8) Frenkel, D. Simulations: The dark side. *Eur. Phys. J. Plus* **2013**, *128*, 10.

(9) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187.

(10) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042.

(11) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

(12) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for ? hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435.

(13) Fieber, W.; Kristjansdottir, S.; Poulsen, P. M. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J. Mol. Biol.* **2004**, *339*, 1191.

(14) Modig, K.; Jürgensen, V. W.; Lindorff-Larsen, K.; Fieber, W.; Bohr, H. G.; Poulsen, F. M. Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett.* **2007**, *581*, 4965.

(15) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **2012**, *51*, 2224.

(16) Camilloni, C.; Schaal, D.; Schweimer, K.; Schwarzinger, S.; De Simone, A. Energy landscape of the prion protein helix 1 probed by metadynamics and NMR. *Biophys. J.* **2012**, *102*, 158.

(17) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058.

(18) Kannan, S.; Zacharias, M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins* **2007**, *66*, 697.

(19) Meli, M.; Colombo, G. A Hamiltonian replica exchange molecular dynamics (MD) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *Int. J. Mol. Sci.* **2013**, *14*, 12157.

(20) Hritz, J.; Oostenbrink, C. Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* **2008**, *128*, 144121.
(21) Kästner, J. Umbrella sampling. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932.
(22) Jiang, W.; Luo, Y.; Maragliano, L.; Roux, B. Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer. *J. Chem. Theory Comput.* **2012**, *8*, 4672.
(23) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553.
(24) Blanco, F. J.; Serrano, L. Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *Eur. J. Biochem.* **1995**, *230*, 634.
(25) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of beta-hairpin formation. *Nature* **1997**, *390*, 196.
(26) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* **2004**, *126*, 7238.
(27) Garcia, A. E; Sanbonmatsu, K. Y. Exploring the energy landscape of a b hairpin in explicit solvent. *Proteins* **2001**, *42*, 345.
(28) Bussi, G.; Gervasio, F. L; Laio, A.l; Parrinello, M. Free-energy landscape for b hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13441.
(29) Bonomi, M.; Branduardi, D.; Gervasio, F. L; Parrinello, M. The unfolded ensemble and folding mechanism of the C-terminal GB1 beta-hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 13938.
(30) Best, R. B.; Mittal, J. Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins* **2011**, *79*, 1318.
(31) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049.
(32) Best, R.; De Sancho, D.; Mittal, J. Residue-specific alpha-helix propensities from molecular simulation. *Biophys. J.* **2012**, *102*, 1462.
(33) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845.
(34) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: a portable plugin for free energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961.
(35) Camilloni, C.; Guerini Rocco, A.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys. J.* **2008**, *94*, 4654.
(36) Pietrucci, F.; Laio, A. A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197.
(37) Patriksson, A.; van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073.
(38) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195.
(39) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011.
(40) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40.
(41) Grossfield, A. *WHAM: an implementation of the weighted histogram analysis method*. http://membrane.urmc.rochester.edu/content/wham (accessed October 14, 2013).
(42) Grossfield, A.l; Feller, S. E.; Pitman, M. C. Convergence of molecular dynamics simulations of membrane proteins. *Proteins: Struct., Funct. Bioinf.* **2007**, *67*, 31.
(43) Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145.
(44) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566.
(45) Berisio, R.; Loguercio, S.; De Simone, A.; Zagari, A.; Vitagliano, L. Polyproline helices in protein structures: A statistical survey. *Protein Pept. Lett.* **2006**, *13*, 847.
(46) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry* **2005**, *44*, 6269.
(47) Berteotti, A.; Barducci, A.; Parrinello, M. Effect of urea on the beta-hairpin conformational ensemble and protein denaturation mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 17200.
(48) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409.
(49) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLoS One* **2012**, *7*, e32131.A.