JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Assessing the Quality of the OPEP Coarse-Grained Force Field

Alessandro Barducci,[†,¶] Massimiliano Bonomi,[†,¶] and Philippe Derreumaux*[,‡]

[†]Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, c/o USI Campus, via Buffi 13, CH-6900 Lugano, Switzerland

[‡]Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique and Université Paris Diderot, Paris 7, Institut Universitaire de France, 13 rue Pierre et Marie Curie, 75005 Paris, France

Supporting Information

ABSTRACT: A coarse-grained potential that could accurately describe the overall conformational landscape of proteins would be extremely valuable not only for structure prediction but also for studying protein dynamics, large conformational motions, and intrinsically disordered systems. Here, we assessed the quality of the OPEP coarse-grained potential by comparing the reconstructed free-energy surfaces (FESs) of two prototypical β-hairpin and α-helix peptides to all-atom calculations in explicit solvent. We found remarkable agreement between the OPEP FES and those obtained using atomistic models, despite a general overstabilization of α- and β-structures by the coarse-grained potential. The use of advanced sampling techniques based on metadynamics and parallel tempering guaranteed a thorough exploration of the conformational space accessible to the two peptides studied.

1. INTRODUCTION

Computer simulations have been proven over the years to be a powerful instrument for getting valuable insight into many biological, physical, and chemical processes. However, many interesting phenomena in these fields of science occur in a time scale—and involve systems with a dimension—that is still not easily treatable in a simulation at atomistic resolution. To bridge the gap between simulation and reality, many different advanced computational techniques have been proposed.[1] Among these is the development of coarse-grained (CG) force fields. This approach consists in representing a configuration of the system in terms of beads at lower resolution and developing an effective interaction between the CG sites that preserves as much as possible the underlying physics. The reduced number of degrees of freedom and a smoother, more-efficient potential allow for systems of bigger size to be simulated for a much longer time scale.

A variety of recipes for coarse-graining a system and building an effective potential have been proposed with the aim of studying protein structure and dynamics, lipid bilayers, nucleic acids, surfactants or polyelectrolytes. Providing an exhaustive review of the many existing CG force fields is beyond the scope of this paper (see refs 2 and 3 for a comprehensive treatment of this topic). In some cases, CG potentials have been parametrized using thermodynamics data from simulations carried out with higher-resolution models, either using force matching or inverse Boltzmann techniques. In the realm of proteins, another class of CG potentials has been developed by fitting analytical functions on a dataset of protein structures resolved by NMR spectroscopy or X-ray crystallography. This type of model can be extremely effective in predicting the native form of those proteins whose three-dimensional structure is unknown. However, since these potentials were parametrized using information about the global free-energy minimum alone, their capability of correctly reproducing the thermodynamics of the overall conformational landscape of a protein is arguable.

A CG potential with such an ability would indeed have a much greater value, because it could be used not only for structure prediction but also for studying protein dynamics, large conformational changes, and systems that are intrinsically disordered. In this respect, a practical way of validating a CG potential is to quantitatively compare it with all-atom (AA) explicit-solvent force fields. In fact, these models have been proven to accurately describe protein conformational landscapes and to reproduce experimental observables.[4,5] In doing this type of comparison, advanced sampling techniques should be used to guarantee a thorough exploration of all the relevant configurations of the proteins chosen as test subjects.

Here, we have focused on OPEP,[6] which is a CG potential that has been developed for proteins. In this model, an amino acid is represented by six beads and the energy function has been fitted to maximize the energy of the native structure and an ensemble of non-native states for a large training set including peptides and proteins.[6,7] OPEP combined with a greedy algorithm and a structural alphabet was able to locate using a benchmark of 25 peptides with 9−23 amino acids, lowest-energy conformations differing by 2.6 A Cα root-mean-square deviation from the full NMR structures.[8] OPEP has also been applied to the amyloid peptide Aβ16-22 and provided structural information on the aggregates consistent with or later confirmed by other computational studies and experimental studies.[9] Finally, OPEP was also able to reproduce the two NMR conformations of the Aβ21-30 peptide in solution.[10]

We assessed the quality of this potential with the help of two peptides that are prototypical examples of β- and α-structures: the C-terminal β-hairpin of protein GB1 and the C-peptide corresponding to the N-terminal residues of RNase A. In particular, we compared the conformational landscape sampled by a

Received: November 10, 2010
Published: May 19, 2011

combined metadynamics and parallel tempering algorithm[11−13] using the OPEP potential and two popular AA force fields in explicit solvent, AMBER99SB[14] and OPLS-AA.[15] For both the peptides, the OPEP force field could properly reproduce the features of the atomistic free-energy landscape, despite a general overstabilization of the β- and α-structures, compared to the AA models.

2. METHODS

2.1. The OPEP Potential. OPEP is based on a six-bead representation of an amino acid, namely, the N, H, Cα, CO, O atoms and one bead or centroid for the side chain. Unlike other CG force fields, which group heavy backbone atoms into interaction centers,[16,17] OPEP is basically an AA backbone with CG side chains. This choice was motivated by a tradeoff between structural resolution, CPU speed, and accuracy of the potential energy function, and, notably, for hydrogen-bond (H-bond) interactions.[6]

The OPEP energy function is defined as a sum of local, nonbonded, and hydrogen-bonded terms. The local term includes the potential energy associated with bond stretching, bending, and torsional angles. These parameters were modeled on the AMBER force field[18] with an additional term in the torsional potential for the Φ and Ψ dihedral angles, which renders realistic Ramachandran plots. The nonbonded potential term is expressed as a sum of van der Waals interactions between a pair of beads. This interaction can be repulsive or attractive, depending on the specific pair of beads. The hydrogen-bond potential has a two-body contribution and a four-body contribution. The latter represents the cooperativity of hydrogen-bond formation and accounts for the propensity of forming different secondary structure as a function of the amino-acid sequence. All the OPEP potential parameters were recently refitted using a training set of 11 protein experimental structures. This version of the potential has been proven to correctly identify 24 native or native-like states for 29 proteins.[7] In this work, we use the OPEP molecular dynamics (MD) code (OPEP-MD), which allows one to efficiently perform both single-replica and replica exchange MD simulations using the OPEP potential.[19,20]

2.2. Metadynamics and Parallel Tempering. Metadynamics is an advanced sampling algorithm that relies on the introduction of a history-dependent potential acting on a selected number of slow degrees of freedom, dubbed collective variables (CVs).[11] If properly applied, metadynamics can both accelerate sampling and reconstruct the free-energy surface (FES) as a function of the CVs. In the well-tempered variant,[21] the FES is obtained by exploiting the relation

$$V(s, t \to \infty) = -\frac{T + \Delta T}{\Delta T} F(s) + C \qquad (1)$$

where s represents the CVs, $V(s,t)$ denotes the bias potential at time t, C is an irrelevant additive constant, T is the temperature of the system, and ΔT denotes an input parameter representing an effective sampling temperature of the CV space.[21] Furthermore, information about the degrees of freedom, other than the CVs, can be easily recovered from a well-tempered metadynamics simulation. In fact, the unbiased probability distribution of a generic function of the microscopic coordinates of the system can be reconstructed using a recently developed reweighting algorithm.[22] This algorithm is particularly helpful whenever one wants to quantitatively compare a simulation with experimental observables.[5]

Table 1. Details of the PTMetaD Simulations of GB1 β-hairpin (HPIN) and RNase C-peptide (CPEP) Using the OPEP CG Potential and the AA Force Fields AMBER99SB and OPLS-AA[a]

	force field	length (ns)	N_{atom}	N_{solv}	N_{rep}	$T_{min}-T_{max}$ (K)
HPIN	AMBER99SB	70	5758	5502	64	270−695
	OPLS-AA	51	5758	5502	64	270−695
	OPEP	660	95	0	16	300−600
CPEP	AMBER99SB	17	3777	3570	64	270−650
	OPLS-AA	22	3777	3570	64	270−650
	OPEP	300	84	0	16	300−600

[a] Details include the length of the PTMetaD run (per replica), the total number of atoms (N_{atom}), the number of solvent atoms (N_{solv}), the number of replicas (N_{rep}), and the temperature range ($T_{min}-T_{max}$).

The capabilities of metadynamics can be further enhanced by combining it with parallel tempering (PT).[12] In the combined method (PTMetaD), several independent metadynamics calculations are performed using the same set of CVs at different temperatures. As in standard PT, configurations are swapped from time to time, following (modified) Metropolis criteria. PTMetaD greatly improves the performance of both PT and single-replica metadynamics.[13]

2.3. Simulation Details. We simulated both the GB1 β-hairpin (GEWTYDDATKTPTVTE) and the RNase C-peptide (AETAAAKFLRNHA) using two AA force-fields (AMBER99SB and OPLS-AA) and the OPEP potential for a total of six independent PTMetaD runs.

In the AA simulations, the system was explicitly solvated using TIP3P water molecules[23] in a rhombic dodecahedron box with periodic boundary conditions. The basic details of the simulations are reported in Table 1. Additional technical details can be found in the Supporting Information.

In the AA PTMetaD runs, exchanges were attempted every 0.2 ps, whereas in the OPEP simulations the stride used was 4 ps. Note that, in all simulations, the average probability of accepting an exchange was ∼0.3−0.4 across the temperature range.

In the case of GB1 β-hairpin, the metadynamics bias was applied to two CVs that describe, respectively, the formation of the hydrophobic core and the population of the backbone−backbone hydrogen bonds associated with the β-sheet formation. In the AA simulations, the first CV was defined as the radius of gyration of all the heavy atoms of the backbone and side chain of residues Trp3, Tyr5, Phe12, and Val14. In the OPEP simulation, we included in the definition the backbone atoms plus the beads that represent the side chains of these residues. The second CV was defined in the same way in both the AA and OPEP simulations as

$$S_\beta = \sum_{i,j} \frac{1 - (d(O_i, HN_j)/d_0)^6}{1 - (d(O_i, HN_j)/d_0)^{12}} \qquad (2)$$

where $d(O_i,HN_j)$ is the distance between the backbone oxygen atom of residue i and the backbone amide hydrogen of residue j, $d_0 = 3$ Å, and the summation runs over all of the pairs of residues whose separation in sequence is greater than 4.

The two metadynamics CVs chosen for the C-peptide were the Cα radius of gyration and the population of the α-helical

hydrogen bonds. This latter CV was defined similarly to S_β, following the example of ref 5:

$$S_\alpha = \sum_{i=1}^{9} \frac{1-(d(O_i,HN_{i+4})/d_0)^6}{1-(d(O_i,HN_{i+4})/d_0)^{12}} \quad (3)$$

where $d(O_i,HN_{i+4})$ is the distance between the backbone oxygen atom of residue i and the backbone amide hydrogen of residue $i+4$ and $d_0 = 3$ Å. Since, for this peptide, both CVs are a function of the backbone atoms alone, the same definition was used, regardless of the force field.

All the AA PTMetaD simulations were performed with GROMACS 4[24] equipped with PLUMED,[25] and metadynamics was coded in OPEP-MD[19] in order to run PTMetaD CG simulations. The metadynamics bias potential was stored on a grid and updated every time a new Gaussian was added, using the implementation available in PLUMED. This computational trick keeps the (additional) cost of metadynamics constant during the simulation and it is particularly convenient whenever the OPEP potential is used. In fact, the cost of computing the CG forces is so small that it would become negligible, compared to the exploding cost of summing, at every MD step, the ever-growing number of Gaussians deposited by metadynamics.

3. RESULTS AND DISCUSSION

3.1. GB1 β-hairpin. The C-terminal domain of the GB1 immunoglobulin binding protein corresponding to residues 41−56 is a prototypical example of β-hairpin structure. It is one of the smallest peptides, displaying two-state behavior with a marginally stable fold at room conditions and a fast folding time of 6 μs.[26] For these reasons, this β-hairpin has attracted the interest of both the experimental community[26−30] and the computational community.[31−53]

Although the details of the folding mechanism are still under debate, it is well-established that folding is a two-state process, in which the turn plays a central role, the secondary structure is formed cooperatively, and both interstrand hydrophobic side-chain−side-chain interactions and backbone hydrogen bonds contribute to the β-hairpin stability. Furthermore, in a recent study,[54] the nature of the configurations belonging to the unfolded ensemble has been investigated and these have been found to be rather compact and dominated by non-native, misfolded β-hairpin structures.

Despite its small size, the conformational ensemble of this peptide appeared to be extremely heterogeneous. Capturing this complexity by projecting the conformations on a low-dimensional set of descriptors is a difficult task. Here, we chose two order parameters that have often been used for studying this β-hairpin:[13,35,41,49] the number of backbone hydrogen bonds and the radius of gyration of few hydrophobic residues (see the Methods section for details). A coordinate similar to the number of hydrogen bonds used here has been recently used to model a stochastic dynamics on it and calculate the folding time distribution of this very same β-hairpin from all-atom replica-exchange simulations.[55]

In Figure 1, we show the FES as a function of these CVs, which were obtained with PTMetaD using the OPEP CG force field and the two AA force fields in explicit solvent (AMBER99SB and OPLS-AA).

The free-energy landscape obtained with the OPEP potential looked qualitatively similar to the AA profiles. The region around five hydrogen bonds and with a radius of gyration equal to 5 Å

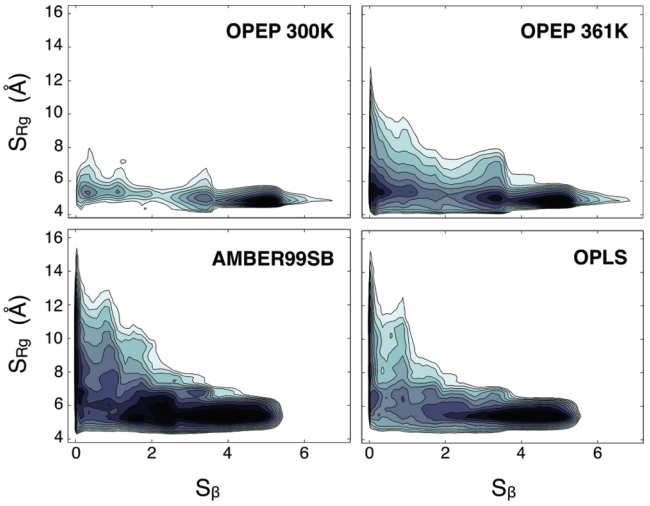

Figure 1. GB1 β-hairpin FES as a function of the number of hydrogen bonds (S_β) and the radius of gyration of the hydrophobic residues (S_{Rg}). The top panels show the OPEP FES at 300 K (left) and 361 K (right), and the bottom panels show the FES from AA simulations in explicit solvent at 300 K using AMBER99SB (left) and OPLS-AA (right) force fields. Isoenergy lines are drawn every 1.5k_BT.

corresponded to the native states of the β-hairpin (see Figure 2, basin A). A cluster analysis of the structures that populate this basin resulted in a single cluster whose central configuration had a root-mean-square deviation (rmsd) of 0.8 Å from the crystallographic structure of the C-terminal of protein GB1.

Basin B contained two main clusters. The former was populated by configurations where the two hydrogen bonds at the termini were broken (Figure 2, basin B_1), the latter by misfolded structures with a shifted pattern of hydrogen bonds (Figure 2, basin B_2). The folded state predicted by the AA potentials more closely resembled the configuration of type B_1. In fact, it is well-known that, in explicit-solvent simulations, the most stable conformation is often a "fried state", in which the terminal hydrogen bonds are not formed. In OPEP, the 4-body term describing the cooperativity of the hydrogen-bond formation and the implicit description of the solvent effects might overstabilize the "full" β-hairpin state.

The unfolded basin corresponded to compact states ($S_{Rg} \simeq 5$ Å) with no hydrogen bonds formed (Figure 2, basin C_1). Part of the structures that belong to this basin presented a single α-helix loop in the central region of the peptide (see Figure 2, basin C_2).

Despite a remarkable, yet qualitative, agreement between the OPEP and the AA FES, it is clear that the native basin in the OPEP landscape at 300 K is overpopulated with respect to the unfolded region. This contrasts both with the recent AA simulations in explicit solvent and with the experimental data which suggested that this peptide was only marginally stable at room temperature.[26,53] To make our comparison more quantitative, we further coarse-grained our description of the β-hairpin by defining those conformations with more than one hydrogen bond formed as being "folded" and those conformations with less than one hydrogen bond formed as being "unfolded". Note that, based on our definition, we included in the "folded" basin not only native-like structures but also misfolded configurations. From the monodimensional FES $F(S_\beta)$, we calculated the free-energy difference ΔF_{FU} between folded (F) and unfolded (U) states for AMBER99SB, OPLS-AA, and OPEP (see the Supporting

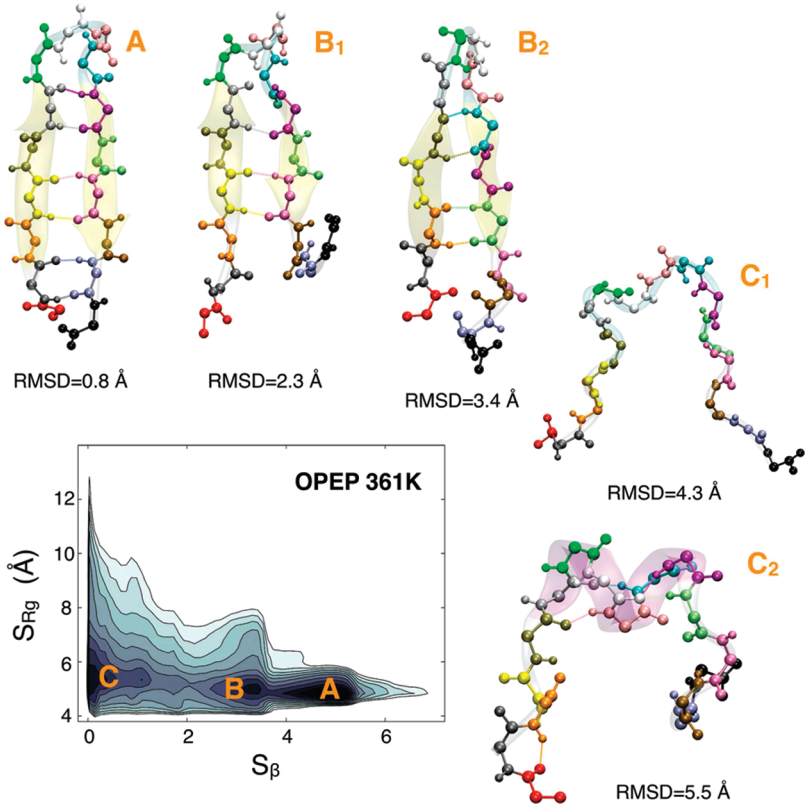

Figure 2. Characterization of the OPEP GB1 β-hairpin FES at 361 K. For each basin, the structure(s) of the most populated cluster(s) is represented. The cluster analysis has been performed using the *g_cluster* tool included in the GROMACS 4 package. The method of Daura and Van Gunsteren[56] was used for clustering, with a cutoff of 1.5 Å on the backbone atoms. The root-mean-sqaure deviation (rmsd) from the crystallographic structure of the C-terminal of protein GB1 (PDB code: 1GB1) was calculated on the backbone.

Table 2. Free-Energy Difference between Folded (F) and Unfolded (U) States of GB1 β-hairpin (ΔF_{FU}) and Unfolding Barrier (ΔF_{TF}) from PTMetaD Calculations Using AMBER99SB, OPLS-AA, and OPEP Force Fields

force field	T (K)	ΔF_{FU} (kcal/mol)	ΔF_{TF} (kcal/mol)
AMBER99SB	300	−2.1	2.5
OPLS-AA	300	−3.4	4.7
OPEP	300	−4.5	7.9
	345	−2.3	6.1
	361	−1.2	5.2

Information for details). The results reported in Table 2 show that the estimates of ΔF_{FU} obtained with two AA models significantly differ and that the AMBER99SB data are more similar to the experimental value.

From $F(S_\beta)$, we also calculated the free-energy difference ΔF_{TF} between the transition state (T) and the folded state (F). This would correspond to the real kinetic barrier that the system must overcome during the unfolding, assuming that S_β is the ideal reaction coordinate of this process. The discrepancy between the values predicted by the two AA models for ΔF_{TF} was even greater than for ΔF_{FU}.

As expected, the overstability observed with the OPEP potential at 300 K could be alleviated by increasing the temperature (see Figure 3). In the temperature range of 345−361 K, the deviation from AA data in terms of both ΔF_{FU} and ΔF_{TF} was minimized.

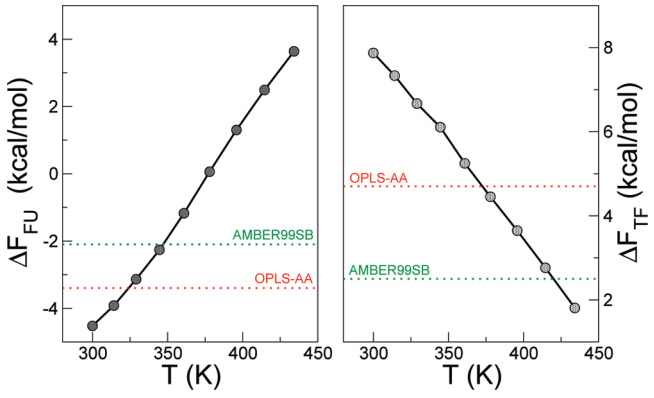

Figure 3. Free-energy difference between folded (F) and unfolded (U) states of GB1 β-hairpin (left panel) and unfolding barrier (right panel), as a function of temperature from the OPEP PTMetaD calculations. Dashed lines indicate the results from the AA simulations at 300 K, using AMBER99SB (green) and OPLS-AA (red).

3.2. RNase C-peptide. The C-peptide corresponding to the first 13 N-terminal residues of Ribonuclease A has a remarkable α-helical propensity for a system of such a small size. In particular, according to circular dichroism experiments, the average helicity is dependent on both temperature and pH and exhibits a maximum at T = 276 K and pH 5.25.[57] NMR experiments[58] and AA MD simulations in explicit solvent on a mutant sequence[5,59] showed that the conformational landscape

of the C-peptide is characterized by an equilibrium among different conformers. These include a set of extended coil conformations, a set of complete α-helical configurations, and a set of partially formed α-helical structures with a salt bridge between the side chains of Glu2 and Arg10.

Here, we reconstructed the FES as a function of the number of α-helical hydrogen bonds and the radius of gyration calculated on the Cα atoms (see the Methods section for details). The FES obtained with PTMetaD, using OPEP, AMBER99SB, and OPLS-AA potentials, are reported in Figure 4.

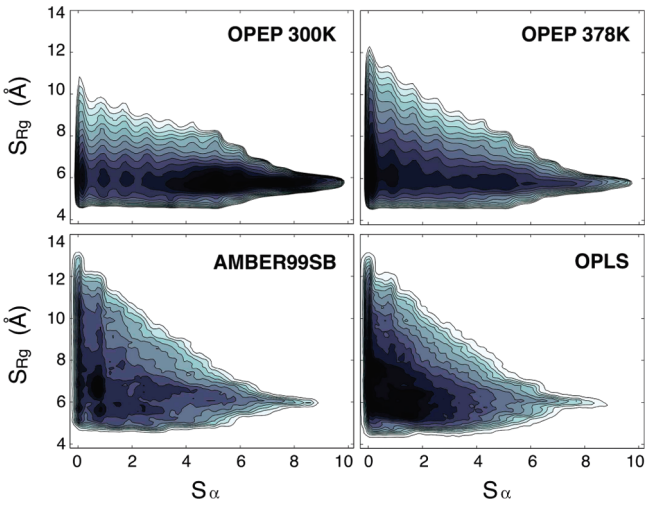

Figure 4. RNase A C-peptide FES, as a function of the number of α-helical hydrogen bonds (S_α) and the Cα radius of gyration (S_{Rg}). The top panels show the OPEP FES at 300 K (left) and 378 K (right), and the bottom panels show the FES from AA simulations in explicit solvent at 300 K using AMBER99SB (left) and OPLS-AA (right) force fields. Isoenergy lines are drawn every $1.5k_BT$.

As for the case of β-hairpin, it is clear from Figure 4 that the OPEP potential overstabilized α-helical conformers at 300 K. This affected both the extended α-helix structure (Figure 5, basin A) and the partially helical conformation that most resembled the crystallographic structure of the N-terminal fragment of Ribonuclease A (Figure 5, basin B).

At higher temperatures, the FES obtained with OPEP looked qualitatively similar to the AA force field landscapes. However, a more quantitative comparison among the FES became problematic, since a two-state (or multistates) behavior was not well-defined as the temperature increased. We decided on a different approach and calculated the propensity of each residue of the C-peptide to form an α-helix. This was defined as the free-energy

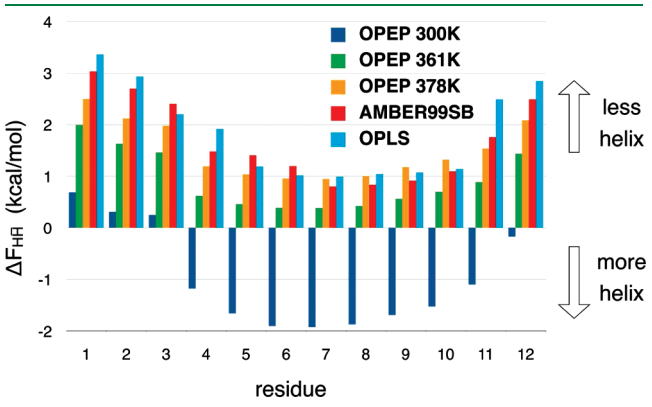

Figure 6. C-peptide helical propensity profile defined as the free-energy difference between helical (H) and nonhelical ($\bar{H}$) states at residue level. The secondary structure was assigned by the STRIDE algorithm. Data were obtained by reweighting the configurations sampled during the PTMetaD run with OPEP at 300 K (blue), 361 K (green), and 378 K (yellow); AMBER99SB at 300 K (red); and OPLS-AA at 300 K (light blue).

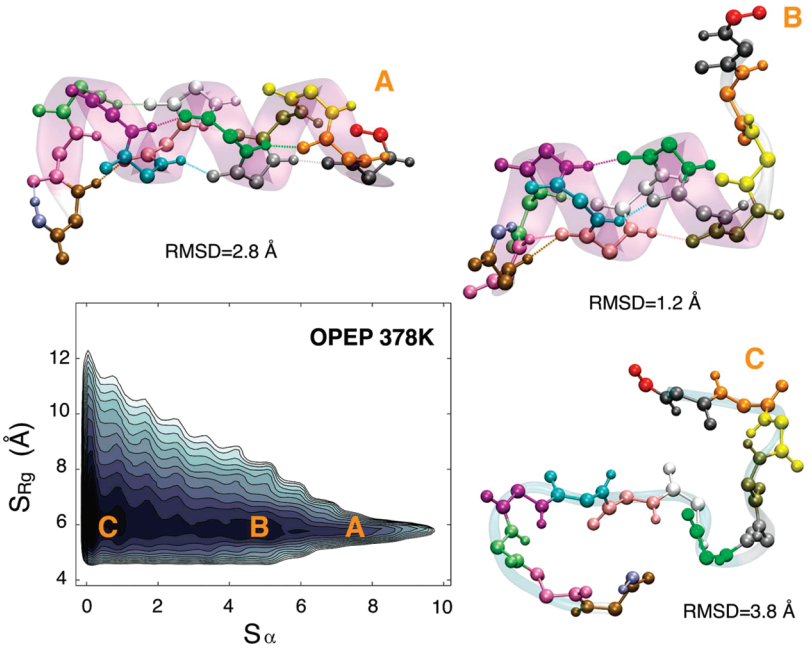

Figure 5. Characterization of the OPEP C-peptide FES at 378 K. The basins that are more clearly recognizable at 300 K become more shallow as temperature increases. (Details of the cluster analysis are reported in the caption of Figure 2.) The rmsd from the crystallographic structure of N-terminal fragment of Ribonuclease A (PDB code: 2AAS) was calculated on the backbone.

Table 3. RMSD of OPEP $\Delta F_{H\bar{H}}$ Data from AA Data at 300 K, as a Function of the OPEP Temperature

T (K)	AMBER99SB (kcal/mol)	OPLS (kcal/mol)
300	2.7	2.9
314	2.3	2.5
329	1.9	2.1
345	1.4	1.6
361	0.8	1.0
378	0.3	0.5
396	0.4	0.5
415	0.7	0.6
434	1.0	0.9

difference $\Delta F_{H\bar{H}}$ between the α-helix state (H) and the states characterized by any other type of secondary structure ($\bar{H}$). We used the STRIDE algorithm[60] to assign the secondary structure to the conformations sampled by the PTMetaD run, and, for each residue, we calculated $\Delta F_{H\bar{H}}$ by reweighting[22] the statistics accumulated during this biased simulation. Results are reported in Figure 6.

At 300 K, the AA force fields predicted a very similar $\Delta F_{H\bar{H}}$ for each residue, with an RMSD of 0.3 kcal/mol between the two sets of data. In contrast, at this temperature, the OPEP potential significantly overstabilized helicity. However, the trend along the peptide sequence was correctly reproduced, with the central residues being the most prone to form α-helix. Also in this case, at higher temperatures, OPEP results became quantitatively comparable to AA data at room temperature (see Table 3). In particular, considering OPEP $\Delta F_{H\bar{H}}$ data at 361 and 378 K, the rmsd values were 1.0 and 0.5 kcal/mol, with respect to OPLS-AA at 300 K, and 0.8 and 0.3 kcal/mol, with respect to AMBER99SB at 300 K.

4. CONCLUSIONS

In this paper, we used an advanced sampling algorithm to describe the conformational landscape of two peptides that are prototypical examples of β-hairpin and α-helical structures. We used three different models for the force field: two with atomistic details (AMBER99SB and OPLS-AA), and one coarse-grained (CG) potential (OPEP). This CG force field was fitted on a dataset of protein structures, in order to maximize the energy of the native configuration and of an ensemble of non-native states.

Our results showed that the OPEP potential could properly reproduce the features of the free-energy landscape obtained with the AA explicit-solvent force fields. However, an overstabilization of both α- and β-structures was observed with OPEP at 300 K. To determine the optimal OPEP temperature that minimized the deviation from AA data at 300 K, we analyzed the temperature dependence of the OPEP folding free-energy difference and unfolding barrier in the case of the β-hairpin and of the α-helical propensity profile in the case of the C-peptide. Our results showed that the optimal temperatures were similar for both peptides and reside in the interval of 345−360 K for the former system and 360−378 K for the latter system. By choosing the OPEP simulation temperature in these regions, the agreement with the AA results at room temperature was of the same quality as that obtained between the two explicit-solvent AA models. Furthermore, it must be pointed out that, if implicit solvation models were added in the comparison, the discrepancies among the AA results would be even greater, at least in the case of the β-hairpin.[61] This information, along with other more-extensive tests, will be used to improve the model in future versions of the OPEP potential.

A key feature of the OPEP potential is the detailed description of the protein backbone and the introduction of specific potential terms accounting for backbone−backbone hydrogen bonding. Several studies showed that the introduction of backbone directional contacts intrinsically stabilized secondary structure elements.[62−64] In order to check the validity of our results in the case of unstructured peptides, we performed OPEP PTMetaD simulation of an intrinsically disordered glycine-serine block copolypeptide $(GS)_8$. This peptide has been recently characterized by means of MD simulations with umbrella sampling, using the OPLS potential in explicit solvent.[65] Even in this case, the results of the OPEP simulations in the temperature range of 360−378 K were comparable with the room-temperature AA data reported in Figure 3 of ref 65 (see the Supporting Information).

In conclusion, here, using advanced sampling techniques, we have shown that simulations carried out with OPEP potential at a universal temperature close to 360 K could properly reproduce the results obtained with atomistic models for a variety of different systems, including ordered and disordered peptides. It must also be noted that, thanks to the reduced representation of the system, OPEP significantly outperforms AA simulations, in terms of computational efficiency. This computational advantage further increases in PT and PTMetaD simulations, because of the fact that fewer replicas are needed to cover the same temperature range.[66] For all these reasons, the combination of the OPEP potential with the PTMetaD sampling algorithm, introduced and tested here for small peptides, paves the way for extending the capabilities of molecular simulations to the study of complex biomolecular processes such as the folding of large proteins and protein−protein interactions.

■ ASSOCIATED CONTENT

Supporting Information. Supporting Information contains additional details of the well-tempered metadynamics algorithm and technical details of the AA and OPEP simulations of the GB1 β-hairpin and RNase C-peptide. General details of the molecular dynamics simulations and specific parameters of the PTMetaD algorithm are given. The convergence of the PTMetaD simulations is discussed in details. The OPEP free-energy landscape of $(GS)_8$, as a function of the radius of gyration, is reported. (PDF) This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: philippe.derreumaux@ibpc.fr.

Author Contributions

¶These authors contributed equally to this work.

■ ACKNOWLEDGMENT

Calculations were performed on the Brutus cluster at ETH Zurich.

REFERENCES

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: New York, 2002; pp 1−617.
(2) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2009.
(3) Tozzini, V. *Q. Rev. Biophys.* **2010**, *43*, 333–371.
(4) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
(5) Barducci, A.; Bonomi, M.; Parrinello, M. *Biophys. J.* **2010**, *98*, L44–L46.
(6) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301–2310.
(7) Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394–408.
(8) Maupetit, J.; Derreumaux, P.; Tuffery, P. *Nucleic Acids Res.* **2009**, *37*, W498–W503.
(9) Santini, A.; Mousseau, N.; Derreumaux, P. *J. Am. Chem. Soc.* **2004**, *126*, 11509–11516.
(10) Chen, W.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2006**, *125*, 084911.
(11) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
(12) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
(13) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
(14) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
(15) Kaminsky, G. A.; T.-R., J.; Friesner, R. A.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
(16) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, P. D.; Marrink, S. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
(17) DeVane, R.; Shinoda, W.; Moore, P. B.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2115–2124.
(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(19) Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2007**, *126*, 025101.
(20) Chebaro, Y.; Dong, X.; Laghaei, R.; Derreumaux, P.; Mousseau, N. *J. Phys. Chem. B* **2009**, *113*, 267–274.
(21) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(22) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.
(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(24) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(25) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
(26) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196–199.
(27) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584–590.
(28) Honda, S.; Kobayashi, N.; Munekata, E. *J. Mol. Biol.* **2000**, *295*, 269–278.
(29) Du, D.; Zhu, Y.; Huang, C.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915–15920.
(30) Muñoz, V.; Ghirlando, R.; Blanco, F. J.; Jas, G. S.; Hofrichter, J.; Eaton, W. A. *Biochemistry* **2006**, *45*, 7023–7035.
(31) Muñoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, E. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872–5879.
(32) Kolinski, A.; Ilkowski, B.; Skolnick, J. *Biophys. J.* **1999**, *77*, 2942–2952.
(33) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U. S.A.* **1999**, *96*, 9068–9073.
(34) Wang, H.; Sung, S. S. *Biopolymers* **1999**, *50*, 763–776.
(35) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062–9067.
(36) Roccatano, D.; Amadei, A.; Nola, A. D.; Berendsen, H. J. C. *Protein Sci.* **1999**, *8*, 2130–2143.
(37) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544–2549.
(38) Ma, B. Y.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091–1104.
(39) Lee, J.; Shin, S. *Biophys. J.* **2001**, *81*, 2507–2516.
(40) García, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345–354.
(41) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.
(42) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12129–12134.
(43) Ma, B.; Nussinov, R. *Protein Sci.* **2003**, *12*, 1882–1893.
(44) Colombo, G.; DeMori, G. M. S.; Roccatano, D. *Protein Sci.* **2003**, *12*, 538–550.
(45) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.
(46) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080–1090.
(47) Wei, G.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56*, 464–474.
(48) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.
(49) Nguyen, P.; Stock, G.; Mittag, E.; Hu, C.; Li, M. S. *Proteins* **2005**, *61*, 795–808.
(50) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 846–859.
(51) Thukral, L.; Smith, J. C.; Daidone, I. *J. Am. Chem. Soc.* **2009**, *131*, 18147–18152.
(52) Juraszek, J.; Bolhuis, P. G. *J. Phys. Chem. B* **2009**, *113*, 16184–16196.
(53) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 8790–8798.
(54) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 13938–13944.
(55) Yang, S.; Onuchic, J. N.; Garcia, A. E.; Levine, H. *J. Mol. Biol.* **2007**, *372*, 756–763.
(56) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
(57) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563–567.
(58) Osterhout, J. J.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1989**, *28*, 7059–7064.
(59) Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. *Biophys. J.* **2005**, *89*, 842–857.
(60) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566–579.
(61) Lwin, T. Z.; Luo, R. *Protein Sci.* **2006**, *15*, 2642–2655.
(62) Hoang, T. X.; Trovato, A.; Seno, F.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U. S.A.* **2004**, *101*, 7960–7964.
(63) Yap, E. H.; Fawzi, N. L.; Head-Gordon, T. *Proteins* **2008**, *70*, 626–638.
(64) Alemani, D.; Collu, F.; Cascella, M.; Peraro, M. D. *J. Chem. Theory Comput.* **2010**, *6*, 315–324.
(65) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2008**, *130*, 7380–7392.
(66) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.