JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

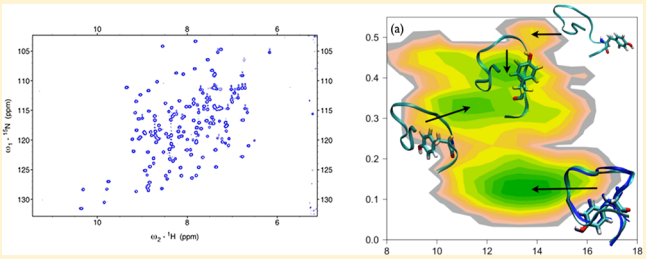
Replica-Averaged Metadynamics

Carlo Camilloni,[†] Andrea Cavalli,[†,‡] and Michele Vendruscolo*[,†]

[†]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW United Kingdom

[‡]Institute for Research in Biomedicine, 6500 Bellinzona, Switzerland

ABSTRACT: A statistical mechanics description of complex molecular systems involves the determination of ensembles of conformations that represent their Boltzmann distributions. The observable properties of these systems can be then predicted by calculating averages over such ensembles. In principle, given accurate energy functions and efficient sampling methods, these ensembles can be generated by molecular dynamics simulations. In practice, however, often the energy functions are known only approximately and the sampling can be carried out only in a limited manner. We describe here a method that enables to increase simultaneously both the quality of the energy functions and of the extent of the sampling in a system-dependent manner. The method is based on the incorporation of experimental data as replica-averaged structural restraints in molecular dynamics simulations and exploits the metadynamics framework to enhance the sampling. The application to the case of α-conotoxin SI, a 13-residue peptide that has been characterized extensively by experimental measurements, shows that the approach that we describe enables accurate free energy landscapes to be generated. The analysis of these landscapes indicates the presence of a low population state in equilibrium with the native state in which the only aromatic residue of α-conotoxin SI is exposed to the solvent, which is a feature that may predispose the peptide to interact with its partners.



INTRODUCTION

Molecular dynamics simulations represent a powerful approach to explore the behavior of molecular systems.[1−4] There are, however, two main limitations in the use of this method to predict the behavior of systems of biological interest: the quality of the energy functions (force fields) and the extent of the sampling of the conformational space, which is associated with the achievable length of the simulations.[1−4] These two issues are not independent, as in order to assess the quality of a given force field one needs to calculate statistical averages of specific properties of a system and compare them with those measured experimentally, for instance, by nuclear magnetic resonance (NMR) spectroscopy, small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), or cryo-electron microscopy (cryo-EM) methods.[5−7] Both an accurate force field and an accurate sampling are therefore needed to establish whether the computations are consistent with the experiments.

In the case of proteins, in recent years, force fields have been improved systematically, both by changing their functional form and by choosing better values for the parameters.[8−12] At the same time, to address the problem of increasing the extent of the sampling, major advances have been made to improve the software and the hardware used to run the simulations[4,13−15] and to develop efficient advanced sampling techniques.[1,2,16] In the latter case, a common feature is that by trading off some information about the dynamics one can sample the relevant conformations by enhancing the crossing of high-energy barriers. Among the most commonly used advanced sampling methods, we mention here those classified as collective variable based techniques (e.g., umbrella sampling[2,17]) or as extended ensemble techniques (e.g., replica-exchange[2,18]). It has also been shown that the incorporation of experimental data as structural restraints in molecular simulations enables the structures and dynamics of proteins to be determined quite accurately.[19−25] Furthermore, it has also been shown that if the restraints are averaged over several copies, or replicas, of a given protein, this approach generates ensembles of conformations compatible with the maximum entropy principle.[26−28] These findings demonstrate that the incorporation of experimental data as replica-averaged structural restrains in molecular dynamics simulations provides an accurate representation of the unknown Boltzmann distribution of a system given an approximate force field and a set of experimental data.[26−28] In practice, however, although the use of replica-averaged structural restraints enables one to improve the quality of the force field for the specific systems for which experimental data are available, it also makes it very challenging to sample the conformational space efficiently, because it increases the number of degrees of freedom proportionally to the number of replicas. In order to be generally applicable, therefore, this approach should be implemented together with enhanced sampling methods.

To address this problem, in this paper we introduce the replica-averaged metadynamics (RAM) method. In this

Received: May 1, 2013
Published: November 5, 2013

   dx.doi.org/10.1021/ct4006272 | J. Chem. Theory Comput. 2013, 9, 5610−5617

approach, experimental data are incorporated as replica-averaged structural restraints in molecular dynamics simulations, and the sampling is enhanced using the metadynamics method, in this case with replica exchange[26−28] in the well-tempered ensemble (WTE) framework.[29,30] This approach improves at the same time, and in a system-dependent manner, the quality of the force field by the direct addition of experimental information and the extent of the sampling by the use of metadynamics. In order to assess the quality of the distribution of structures obtained by the RAM method, we compare the results obtained with it with those obtained by sampling the unmodified force field through the use of the parallel-tempering metadynamics (PTMetaD) technique.[31,32] In this sense, our goal here is not to compare the efficiency of the sampling of the RAM and PTMetaD methods, but instead to identify the differences in the free energy landscapes associated with the incorporation of experimental information to generate replica-averaged structural restraints in the simulations. The PTMetaD and RAM methods are represented schematically in Figure 1.

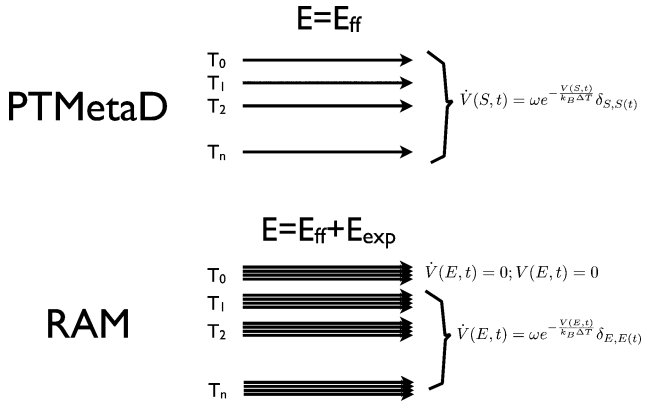


Figure 1. Schematic representation of the PTMetaD and RAM methods. In the PTMetaD method, metadynamics, in the version of the RAM method discussed in this work, is used together with parallel tempering to enhance the sampling of the unmodified force field. By contrast, in the RAM method the experimental data are used as replica-averaged structural restraints to enhance on the fly in a system-dependent manner the quality of the force-field, while metadynamics is used to increase the acceptance rate between replicas at different temperature in the WTE approach[29,30] to parallel tempering.

In order to illustrate the RAM method, we studied α-conotoxin SI,[33] a neurotoxic 13-residue peptide with two disulfide bridges that strongly constraint the accessible conformational space of this molecule (Figure 2a). The small dimensions of this system, with a well-defined folded conformation and wide range of experimental data available (including a high-resolution X-ray structure, NMR chemical shifts, $^3J_{HNHA}$-couplings, and NOE-derived distances), make it well suited for the purpose of the present work. Furthermore α-conotoxins bind selectively in a way that is not yet fully understood nicotinic acetylcholine receptors, which represent a diverse family of homo- or heteropentameric ligand-gated ion channels.[34] A detailed characterization of the conformational space populated by this peptide is thus expected to be helpful for future studies of the mechanism of binding between α-conotoxins and its associated receptors.

METHODS

The simulations described here were performed using the Amber03W force field[35,36] with the TIP4P05 water model.[37] All the simulations were run in GROMACS[14] modified with PLUMED[38] and Almost.[39] A time step of 2 fs was used together with LINCS constraints.[40] The van der Waals interactions were implemented with a cutoff at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method.[41] All simulations were carried out in the canonical ensemble by keeping the volume fixed and by thermosetting the system with the Nosè−Hoover thermostat.[42] The starting conformation was taken from an available X-ray structure (PDB code 1HJE). This structure was protonated and solvated with 1224 water molecules in a dodecahedron box of 43 nm^3 of volume. The energy of the system was first minimized and then the temperature was increased to 283 K in two separate steps, in the first one a 50 ps simulation was performed by keeping fixed the heavy atoms of the protein, and successively a second 200 ps simulation was performed without restraints. The density of the system was relaxed by a 200 ps run using the Berendsen barostat.[43]

In the RAM simulations, the experimental data are used to modify on the fly in a system-dependent manner the underlying force field by employing a replica averaging procedure over back-calculated experimental parameters, which are compared with their experimentally measured values. In this way, in the implementation of the RAM approach discussed here, the ensemble of structures resulting from the simulations at the lowest temperature represents the free energy of the force field modified, in this case, by NMR chemical shifts. The equilibrium values of the observables can be thus calculated directly on this ensemble without the use of a reweighting procedure.[26−28] In the implementation of the RAM method described here, we augmented the force field using a term based on the replica averaging of the backbone NMR chemical shifts,[44] where the chemical shifts themselves are back-calculated at each time step and averaged over four replicas of the system.[45] The replica-averaged chemical shifts are restrained to be close to their experimental values using a quadratic potential.[45] As a parallel tempering approach applied to the RAM case would require four times more replicas than in a standard parallel tempering, to reduce the number of replicas we employed the WTE technique, which is a metadynamics approach that uses the energy as the only collective variable.[29,30] The six temperatures used were 283, 288, 318, 353, 394.2, and 443.1 K, so that also in this case 24 replicas were needed. Furthermore, the metadynamics bias was not used over the first four replicas, in such a way that the resulting ensemble was biased only by the chemical shifts. In this way an ensemble of structures that incorporates the information provided by the experimental chemical shifts in the sense of the maximum entropy principle[26−28] can be easily obtained and analyzed. Alternatively, the RAM method could be implemented by using the metadynamics also on the first four replicas, adopting then a reweighting procedure to obtain the correct statistical weights of the conformations. The bias factor of the well-tempering was set to 10, and the exchanges were tried every 250 steps. Each replica was simulated for 60 ns, where the first 10 ns were discarded, and the convergence was tested by comparing two halves of the simulation (Figure 2b).

In the PTMetaD simulations, the parallel tempering method, in which multiple copies of a system are simulated at different

dx.doi.org/10.1021/ct4006272 | J. Chem. Theory Comput. 2013, 9, 5610−5617

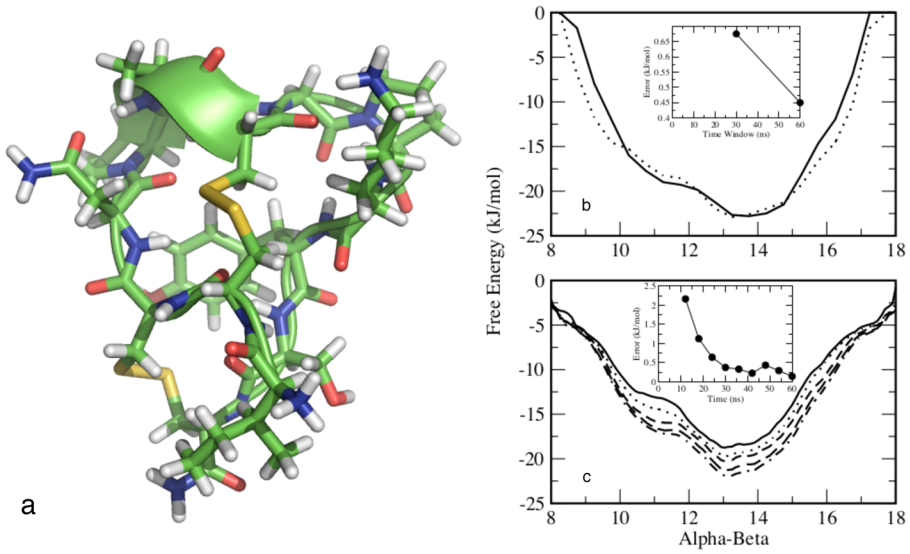


Figure 2. (a) X-ray structure of α-conotoxin SI (PDB code 1HJE). The two disulfide bonds are shown in yellow. (b) Comparison of the free energy profiles calculated for the Alpha-Beta collective variable for the first and second halves of the RAM ensemble. (c) Comparison of free energy profiles for the PTMetaD calculations plotted at 6 ns intervals from 36 to 60 ns. In both simulations the round trip time for the replica-exchange is estimated to be of around 6 ns. These results illustrate the convergence of the PTMetaD and RAM simulations. The insets in panels b and c indicate that the errors in the estimates of the free energies are small compared to the thermal energy of the system.

temperatures and exchanges between neighboring simulations are performed with a Metropolis criterion, is coupled with the metadynamics method on a set of collective variables.[31,32] The coupling of the two methods ensures that slow degrees of freedom not included in the collective variables are sampled correctly through parallel tempering. In the PTMetaD simulations that we carried out here, we used two collective variables. The first, called CamShift,[46] is the root-mean-square difference between calculated and experimental (BMRB 4503)[33] chemical shifts, where the calculated chemical shifts are obtained using the CamShift method;[47] for this collective variable, we used a functional form described previously.[48] The second collective variable, called Alpha-Beta, is a measure of the deviation from zero of the φ and ψ dihedral angles as implemented in PLUMED[38] and defined as

$$AB = \frac{1}{4}\sum_{Nres}^{i}[(1+\cos(\varphi_i)) + (1+\cos(\psi_i))]$$

where the sum runs over all the N_{res} residues of the protein. The temperatures, which were chosen using an algorithm proposed recently,[49] were 283, 288, 293.1, 298.4, 303.8, 309.4, 315.1, 321.0, 327.0, 333.3, 339.7, 346.2, 353.0, 360.0, 367.2, 374.6, 382.2, 390.1, 398.2, 406.6, 415.2, 424.2, 433.4, and 442.9 K for a total of 24 replicas. The Gaussian functions in the metadynamics were deposited every 200 steps, with a height of 0.05 kJ/mol and a width of 5.0 and 0.1, respectively, for the CamShift and Alpha-Beta collective variables;[46] a bias factor of 10 was used to rescale the Gaussian functions as in the well-tempered formulation of metadynamics.[50] Exchanges were tried every 250 steps. In order to back-calculate collective variables and observables not included in the metadynamics simulations, we used a reweighting algorithm.[51−53] Each individual replica was simulated for 60 ns, which we verified was sufficient to reach convergence in the free energy calculations (Figure 2c).

The exchange rules for the PTMetaD and the RAM methods follow those of the replica-exchange method.[44,45] In the replica-exchange approach, multiple replicas of a system are simulated in parallel and exchanges between the replicas are tried at a given rate and accepted with a probability p according to the Metropolis rule[2], $p = \min(1,\exp(-\Delta))$. In the PTMetaD method,[31] independently from the collective variables employed, given two replicas i and j the factor Δ required to calculate the probability of exchange is

$$\Delta_{ij}^{\mathrm{PTMetaD}} = (\beta_i - \beta_j)(E(X_j) - E(X_i)) - \beta_i(V_i(q(X_i)) - V_i(q(X_j))) - \beta_j(V_j(q(X_j)) - V_j(q(X_i)))$$

where β is the inverse of the temperature times the Boltzmann constant, E is the value of the force field for a configuration X, and $V(q(X))$ is the value of the bias resulting from the metadynamics rules on the collective variable q. In the version of the RAM method discussed here, at each temperature β, M replicas (four in this case) are coupled together by a restraint imposed on the average value of multiple observables.[26−28] Furthermore, all the replicas of each block b, except those at the reference temperature ($b = 0$), experience also a metadynamics bias on the energy.[51] In principle, the exchange rule of two replicas belonging to two different blocks should include the back-calculation of the new average chemical shifts. A problem with this approach, however, is that the coupling of the replicas resulting from the chemical shifts reduces the exchange probability. Furthermore, one can only try one exchange between two blocks at different temperatures, thus resulting in a poor exchange rate. Alternatively, one can attempt to exchange all the replicas at two given temperatures. This second approach does not need the recalculation of the chemical shifts, and a move is accepted using the product of the four probabilities of exchange. In the present implementation, we used this second exchange move, and because of the enhancement of the probability of exchange resulting from the WTE approach, we obtained an average acceptance rate of the 10%. Concerning the factor Δ_{ij}, for two replicas belonging to two different blocks, this factor is the same of PTMetaD with

dx.doi.org/10.1021/ct4006272 | J. Chem. Theory Comput. 2013, 9, 5610−5617

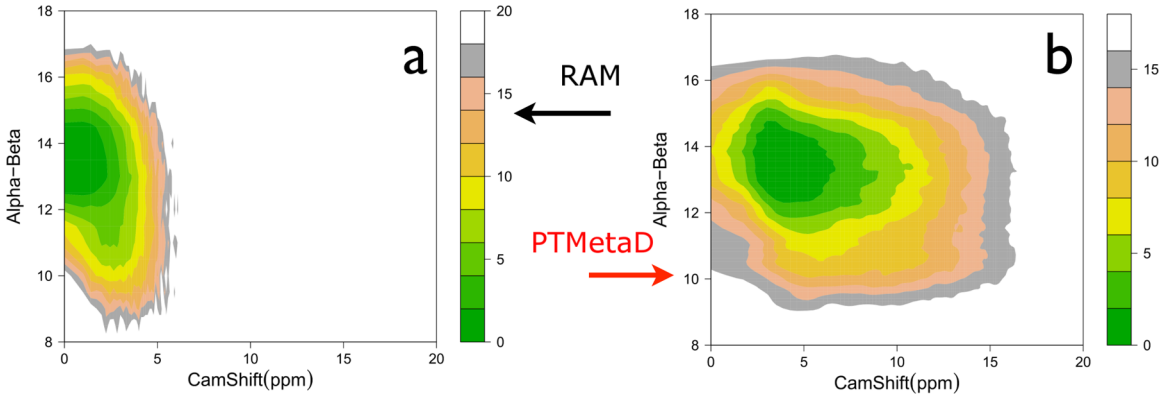


Figure 3. Comparison of the free energy landscapes of the RAM (a) and PTMetaD (b) ensembles as a function of the collective variables utilized in the PTMetaD simulations (on the *x*-axis we report the square root of the CamShift collective variable). The isolines are plotted at 2 kJ/mol intervals. In the PTMetaD simulations, the ensemble spans a larger range of the collective variable CamShift (see Methods), while both the ensembles cover a similar range of value for the collective variable Alpha-Beta (see Methods).

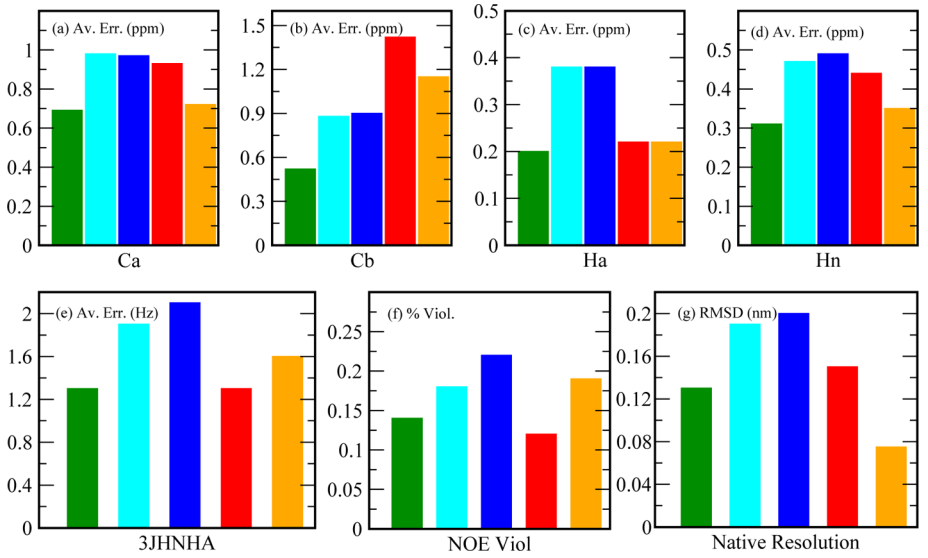


Figure 4. Assessment of the quality of the different structural ensembles of α-conotoxin SI described in this work (see also Table 1). The RAM ensemble is shown in green, the PTMETAD ensemble in cyan, the MD ensemble in blue, the NMR structure[33] (PDB code 1QMW, which was generated by using NOEs and $^3J_{HNHA}$-couplings) in red, and the X-ray structure (PDB code 1HJE) in orange. (a−d) Comparison of the errors in the chemical shifts of Cα, Cβ, Hα, and H_N atoms, respectively. (e) Comparison of the errors in the $^3J_{HNHA}$-couplings. (f) Comparison of the NOE violations. (g) The native resolution is defined as the RMSD with respect to the X-ray structure calculated over all the atoms of a representative structure of the major cluster.

the difference that now $E(X)$ that is the energy of the force-field includes also the energy of the chemical shift restraints

$$E_i^{cs}(X_i) = \alpha_i \sum_{N}^{k=1} \sum_{6}^{l=1} \left(\delta_{kl}^{exp} - \frac{1}{M} \sum_{M}^{m=1} \delta_{kl}^{calc}(X_m) \right)^2$$

where the α_i are force constants, k runs over the amino acids of the peptide or protein, l runs over the six backbone atoms used in the simulations (Cα, Cβ, C′, Hα, H_N, and N), and m runs over the M replicas composing the ith block.

RESULTS

In order to assess the quality of the ensembles obtained by the RAM method, we applied it to determine the conformational fluctuations in the native state of α-conotoxin SI, and compared the resulting free energy landscape with the corresponding one obtained for an unmodified force field using the PTMetaD method (Figure 3). Since in both cases we reached convergence, the differences in the free energy landscapes are a consequence of the additional restraining term. Both simulations explored a similar range of values for the collective variable Alpha-Beta (see Methods) of the backbone, while the PTMetaD explored a wider range of values of the collective variable CamShift (see Methods), which correspond to conformations with larger deviations from the experimental chemical shifts. In the PTMetaD simulations, the extension of the lower region of the free energy with respect to CamShift is larger than in the RAM case, indicating that structures with very different levels of agreement with the experimental data have similar statistical weights.

The quality of the two extensive samplings of the conformational space that we have described above has been assessed using different sets of experimental data. In the RAM ensemble, after discarding the first 10 ns, all the structures sampled by the four replicas at the lowest temperature were used for the analysis. In the PTMetaD ensemble all the

dx.doi.org/10.1021/ct4006272 | J. Chem. Theory Comput. 2013, 9, 5610−5617

Table 1. Assessment of the Quality of the Different Structural Ensembles of α-Conotoxin SI Described in This Work[a]

experiment	RAM	PTMetaD	RA	MD	NMR (1QMW)	X-ray (1HJE)
Cα (ppm)	**0.69**	0.98	0.72	0.97	0.93	0.72
Cβ (ppm)	**0.52**	0.88	0.59	0.90	1.42	1.15
Hα (ppm)	**0.20**	0.38	0.23	0.38	0.22	0.22
H_N (ppm)	**0.31**	0.47	0.32	0.49	0.44	0.35
$^3J_{HNHA}$ (Hz)	**1.30**	1.90	1.51	2.10	**1.30**	1.60
NOE violations	20(10) /143	26(18) /143	23(12) /143	32(21) /143	**17(2) /143**	27(13) /143

[a]Average differences between the experimental values and those back-calculated from the different ensembles. The best agreement between experimental and calculated parameters is shown in bold. The RAM and PTMetaD ensembles are discussed in the main text, while the molecular dynamics (MD) ensemble is generated with a NVT 250 ns simulation; this simulation never leaves the native minimum of the PTMetaD ensemble. The higher quality of the PTMetaD ensemble with respect to the MD ensemble illustrates the importance of taking account of a the broad ensemble of conformations explored by a protein through its equilibrium fluctuations when calculating NMR observables; this observation holds in particular for the Cβ chemical shifts, which are related to the averaging of the side chains. The higher quality of the RAM ensemble with respect to the PTMetaD ensemble shows how the use of the experimental data as replica-averaged structural restraints can improve in a system-dependent manner the underlying force field. The comparison of the NMR structure[33] (PDB code 1QMW, which was generated by using NOEs and $^3J_{HNHA}$-couplings) and the X-ray structure (PDB code 1HJE) shows that a single structure cannot fully reproduce all the available experimental data. For the NOE violations, the numbers in parentheses indicate the NOE violations larger than 0.05 nm.

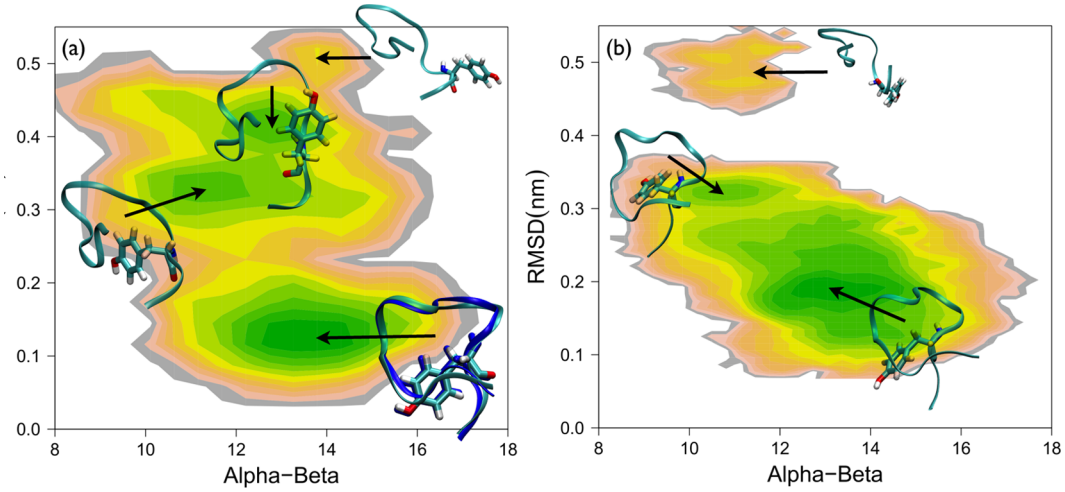


Figure 5. Comparison of the free energy landscapes of the RAM (panel a) and PTMetaD (panel b) ensembles as a function of the collective variable Alpha-Beta (see Methods) and the all-atom RMSD with respect to the X-ray structure (isolines at 2 kJ/mol). The ensembles span a similar range of values, but there are notable differences. The RAM ensemble shows a native minimum closer to the X-ray structure than that of the PTMetaD ensemble (the RMSD value is 0.1 nm for the RAM simulations while it is 0.2 nm for the PTMetaD simulations). The X-ray structure is shown in blue in the lower right corner of panel superposed to the central structure of most populated cluster (a). A second minimum is almost in the same place for both the ensembles, with a RMSD value of 0.3 nm and an Alpha-Beta value of 11. A third minimum at a RMSD value of 0.4 nm and an Alpha-Beta value of 13.5 are only present in the RAM ensemble, and a fourth minimum at a RMSD value of 0.5 nm is present in both the ensembles but with different Alpha-Beta values (14 for the RAM and 11 for the PTMetaD simulations).

structures sampled by the lowest temperature replica have been used with a standard reweighing procedure.[51] Chemical shifts were back-calculated using the SPARTA+ method,[54] $^3J_{HNHA}$-couplings were back calculated using the Karplus equations with the "DFT2" set of parameters[55] and NOE distances were back calculated by using a $1/r^6$ distance dependence.[21] The average errors of the back-calculated data with respect to the experimental sets are reported in Figure 4 and Table 1. In all cases, the RAM ensemble was found to be in better agreement with the experimental data than the PTMetaD ensemble (Figure 4 and Table 1), thus indicating that the use of replica-averaged chemical shift restraints enables to find an ensemble of structures in better agreement with experimental observations.

In order to test whether the Amber03W force field corresponds to a well-defined folded conformation, we used the PTMetaD free energy landscape, which was generated without the use of structural restraints, as a function of the collective variable Alpha-Beta and the RMSD calculated over all the atoms with respect to an X-ray structure (1HJE, Figure 5). We found that in the PTMetaD ensemble the native minimum is farther from the X-ray structure (0.18 nm) than the RAM ensemble (0.11 nm). A second minimum is almost in the same place for both the ensembles, with a RMSD value of 0.3 nm and an Alpha-Beta value of 11. A third minimum at a RMSD value of 0.4 nm and an Alpha-Beta value of 13.5 is only present in the RAM ensemble, and a fourth minimum at a RMSD value of 0.5 nm is present in both the ensembles but with a different value of the Alpha-Beta collective variable (14 for the RAM and 11 for the PTMetaD ensembles).

The native state population has been estimated at 92%,[33] a value that corresponds quite accurately to the value of 90% that we obtain for the RAM ensemble by considering all the conformations in which Tyr11 is buried (first and second minima in Figure 5a). Representative conformations for these minima are shown in Figure 6a,b. Further analysis of the free energy landscapes reveals the presence of low population states

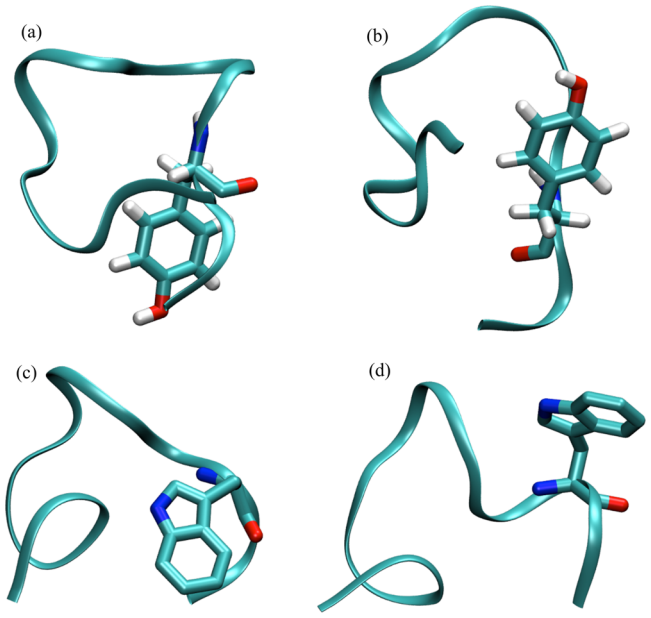


Figure 6. Comparison of the structures corresponding to the major (panel a) and minor (panel b) populations of α-conotoxin SI, and the free (1CNL, panel c) and bound (2C9T, panel d) states of α-conotoxin IMI. In the major population of α-conotoxin SI, the only aromatic residue (Tyr11) or this peptide is buried, and correspondingly Trp10 is buried in the free state of α-conotoxin IMI. In the minor population of α-conotoxin SI, Tyr11 is exposed, and correspondingly, Trp10 points outward in the bound state of α-conotoxin IMI.

(third and fourth minima in Figure 5a) that differ from the native state primarily in terms of the orientation of a tyrosine residue at position 11 along the sequence (Tyr11), which is buried in the structure in the native state but exposed to the solvent in the low population states. Although the major effect of the structural restraints that we incorporated in the simulations should be expected to be reflected in the quality of the structures corresponding to the most populated minima, as higher-energy structure would contribute much less to the statistical averages, these results suggest the intriguing possibility that the exposure of Tyr11, which is the only aromatic residue of α-conotoxin SI (sequence ICCNPACGPKYSC),[33] in the low population state (Figure 6b) may create the possibility of forming intermolecular interactions. Although there are no structures in the PBD of α-conotoxin SI in a bound state, we investigated this idea by analyzing the bound state of α-conotoxin IMI, a peptide closely homologous to of α-conotoxin SI, in which Tyr11 is replaced by another aromatic residue, Trp10 (Figure 6c). In this case of α-conotoxin IMI, Trp10 is oriented in the bound state so that its side chain points outward in a configuration that stabilizes the complex (Figure 6d, where only α-conotoxin IMI is shown in the bound state). Based on these results, we can speculate that the exposure of Tyr11 in the low population state may predispose α-conotoxin SI to bind its partners.

These results show that the addition of the chemical shifts as replica-averaged structural restraints in the simulations, by changing the free energy landscape of the system to increase the agreement with the experimental data, increases the overall quality of the resulting conformations. The two ensembles generated by the RAM and PTMetaD methods can also be compared with the experimental structures and with both a standard molecular dynamics simulations and a standard Replica-Averaged simulations.[45] Both the 250 ns unbiased molecular dynamics simulation (MD run) and four replica-averaged (RA, i.e. the RAM method without metadynamics) 60 ns long molecular dynamics simulations are in an overall worse agreement with all the available data than the corresponding simulations in which the sampling is enhanced (PTMetaD and RAM), thus showing the importance of enhanced sampling techniques in exploring high energy conformations and that of replica-averaging in correcting the force-field. RAM can address both the problems at the same time. By contrast, a previously determined NMR structure[33] (PDB code 1QMW) was found to be in better agreement with NOEs and $^{3}J_{HNHA}$-couplings, which were the data employed to generate it,[33] than the PTMetaD ensemble; the agreement with the chemical shifts is poorer for the 1QWM structure than the PTMetaD ensemble. The X-ray structure too is not in good agreement with the chemical shifts and the $^{3}J_{HNHA}$-couplings, despite having only a few NOE violations. These results show the importance of considering conformational averaging in the comparison with experimental measurements also in the case of a peptide with a well-defined folded conformation and suggest that the RAM method will be useful in the case of disordered system such as intrinsically disordered proteins.

CONCLUSIONS

We have described the RAM method, which combines the use of experimental data as replica-averaged structural restraints in molecular dynamics simulations and the metadynamics scheme. This approach enables one to simultaneously enhance the sampling efficiency and improve the force field for specific systems, and thus represents an efficient tool to generate ensemble of structure representing the structure and the dynamics of proteins.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mv245@cam.ac.uk .

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.C. was supported by a Marie Curie Intra European Fellowship.

REFERENCES

(1) Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52−61.
(2) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: London, 2002.
(3) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679−6685.
(4) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341−346.
(5) Arkhipov, A.; Shan, Y. B.; Das, R.; Endres, N. F.; Eastwood, M. P.; Wemmer, D. E.; Kuriyan, J.; Shaw, D. E. Architecture and membrane interactions of the EGF receptor. *Cell* **2013**, *152*, 557−569.
(6) Cellmer, T.; Buscaglia, M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. Making connections between ultrafast protein folding kinetics and molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6103−6108.
(7) Ward, A. B.; Sali, A.; Wilson, I. A. Integrative structural biology. *Science* **2013**, *339*, 913−915.

(8) Best, R. B.; Buchete, N. V.; Hummer, G. Are current molecular dynamics force fields too helical? *Bioph. J.* **2008**, *95*, L7−L9.
(9) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ, and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theor. Comput.* **2012**, *8*, 3257−3273.
(10) Li, D. W.; Bruschweiler, R. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. *J. Chem. Theor. Comput.* **2011**, *7*, 1773−1782.
(11) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Bioph. J.* **2011**, *100*, L47−L49.
(12) Stone, A. J. Intermolecular potentials. *Science* **2008**, *321*, 787−789.
(13) Freddolino, P. L.; Harrison, C. B.; Liu, Y. X.; Schulten, K. Challenges in protein-folding simulations. *Nat. Phys.* **2010**, *6*, 751−758.
(14) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comput.* **2008**, *4*, 435−447.
(15) Shirts, M.; Pande, V. S. Computing—Screen savers of the world unite! *Science* **2000**, *290*, 1903−1904.
(16) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(17) Torrie, G. M.; Valleau, J. P. Non-physical sampling distributions in monte-carlo free-energy estimation−umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(18) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(19) Bonvin, A.; Boelens, R.; Kaptein, R. Time-averaged and ensemble-averaged direct NOE restraints. *J. Biomol. NMR* **1994**, *4*, 143−149.
(20) Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. Conformational dynamics detected by nuclear magnetic-resonance NOE values and J-coupling constants. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396.
(21) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005**, *433*, 128−132.
(22) Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A. Atomic model of the type III secretion system needle. *Nature* **2012**, *486*, 276−279.
(23) Pieper, U.; Schlessinger, A.; Kloppmann, E.; Chang, G. A.; Chou, J. J.; Dumont, M. E.; Fox, B. G.; Fromme, P.; Hendrickson, W. A.; Malkowski, M. G.; Rees, D. C.; Stokes, D. L.; Stowell, M. H. B.; Wiener, M. C.; Rost, B.; Stroud, R. M.; Stevens, R. C.; Sali, A. Coordinating the impact of structural genomics on the human α-helical transmembrane proteome. *Nat. Struct. Mol. Biol.* **2013**, *20*, 135−138.
(24) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time-dependent distance restraints in molecular-dynamics simulations. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(25) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001**, *409*, 641−645.
(26) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, 094112.
(27) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theor. Comput.* **2012**, *8*, 3445−3451.
(28) Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, 084107.
(29) Bonomi, M.; Parrinello, M. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **2010**, *104*.
(30) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theor. Comput.* **2012**, *8*, 2189−2192.
(31) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(32) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(33) Benie, A. J.; Whitford, D.; Hargittai, B.; Barany, G.; Janes, R. W. Solution structure of α-conotoxin si. *FEBS Lett.* **2000**, *476*, 287−295.
(34) Lee, C.; Lee, S. H.; Kim, D. H.; Han, K. H. Molecular docking study on the α 3 beta 2 neuronal nicotinic acetylcholine receptor complexed with α-conotoxin GIC. *BMB Rep.* **2012**, *45*, 275−280.
(35) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.
(36) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(37) Abascal, J. L. F.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*.
(38) 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*, 1961−1972.
(39) Cavalli, A.; Vendruscolo, M.; Paci, E. Comparison of sequence-based and structure-based energy functions for the reversible folding of a peptide. *Bioph. J.* **2005**, *88*, 3158−3166.
(40) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Comput.* **2008**, *4*, 116−122.
(41) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(42) Evans, D. J.; Holian, B. L. The Nose-Hoover thermostat. *J. Chem. Phys.* **1985**, *83*, 4069−4074.
(43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(44) Camilloni, C.; Robustelli, P.; De Simone, A.; Cavalli, A.; Vendruscolo, M. Characterization of the conformational equilibrium between the two major substates of RNase a using NMR chemical shifts. *J. Am. Chem. Soc.* **2012**, *134*, 3968−3971.
(45) Camilloni, C.; Cavalli, A.; Vendruscolo, M. Assessment of the use chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterise the dynamics of proteins. *J. Phys. Chem. B* **2013**, *117*, 1838−1843.
(46) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterisation of the free energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6817−6822.
(47) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* **2009**, *131*, 13894−13895.
(48) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **2010**, *18*, 923−933.
(49) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations. *J. Chem. Theor. Comput.* **2011**, *7*, 2025−2027.
(50) Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*.

(51) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(52) Camilloni, C.; Schaal, D.; Schweimer, K.; Schwarzinger, S.; De Simone, A. Energy landscape of the prion protein helix 1 probed by metadynamics and NMR. *Bioph. J.* **2012**, *102*, 158−167.
(53) Barducci, A.; Bonomi, M.; Parrinello, M. Linking well-tempered metadynamics simulations with experiments. *Bioph. J.* **2010**, *98*, L44−L46.
(54) Shen, Y.; Bax, A. SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **2010**, *48*, 13−22.
(55) Case, D. A.; Scheurer, C.; Bruschweiler, R. Static and dynamic effects on vicinal scalar J couplings in proteins and peptides: A MD/DFT analysis. *J. Am. Chem. Soc.* **2000**, *122*, 10390−10397.