Communication: Role of explicit water models in the helix folding/unfolding processes

Ferruccio Palazzesi' , Matteo Salvalaglio, Alessandro Barducci, and Michele Parrinello

Citation: J. Chem. Phys. **145**, 121101 (2016); doi: 10.1063/1.4963340
View online: http://dx.doi.org/10.1063/1.4963340
View Table of Contents: http://aip.scitation.org/toc/jcp/145/12
Published by the American Institute of Physics

Communication: Role of explicit water models in the helix folding/unfolding processes

Ferruccio Palazzesi,[1,2,a)] Matteo Salvalaglio,[3] Alessandro Barducci,[4,5] and Michele Parrinello[1,2]

[1]*Department of Chemistry and Applied Biosciences, Eidgenössische Technische Hochschule Zürich, 8093 Zurich, Switzerland*
[2]*Facoltá di Informatica, Istituto di Scienze Computazionali, Universitá della Svizzera Italiana, 6900 Lugano, Switzerland*
[3]*Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom*
[4]*Inserm, U1054 Montpellier, France*
[5]*Université de Montpellier, CNRS, UMR 5048, Centre de Biochimie Structurale, Montpellier, France*

(Received 11 July 2016; accepted 13 September 2016; published online 26 September 2016)

In the last years, it has become evident that computer simulations can assume a relevant role in modelling protein dynamical motions for their ability to provide a full atomistic image of the processes under investigation. The ability of the current protein force-fields in reproducing the correct thermodynamics and kinetics systems behaviour is thus an essential ingredient to improve our understanding of many relevant biological functionalities. In this work, employing the last developments of the metadynamics framework, we compare the ability of state-of-the-art all-atom empirical functions and water models to consistently reproduce the folding and unfolding of a helix turn motif in a model peptide. This theoretical study puts in evidence that the choice of the water models can influence the thermodynamic and the kinetics of the system under investigation, and for this reason cannot be considered trivial. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4963340]

I. INTRODUCTION

Secondary structures are local patterns defining the three-dimensional arrangement of protein segments. Mechanisms underlying the folding and unfolding of such motifs are of uttermost importance in biology, and have been extensively investigated in the scientific literature.[1–4] Driven by the desire to obtain a microscopic detailed picture of these dynamical mechanisms and thanks to progress in algorithms and hardware,[5–7] molecular dynamics (MD) simulations have rapidly become an important investigation tool in protein folding.[8–11] Nevertheless, as several authors have shown,[12–21] the results of MD calculations may be substantially affected by the choice of the empirical protein force-field (PFF) and of the water model (WM).

Here, we provide a detailed microscopic picture of the folding and unfolding of a small helix turn motif, with the aim of assessing the impact of state-of-the-art all-atom empirical potentials on the description of these dynamical processes.

With this aim in mind, we focus our efforts on the formation and disruption of a helical nucleus in a capped five-residues peptide, Ac-WAAAH$^+$-NH$_2$ (WH5). Recent experimental and computational studies have demonstrated that this peptide displays a very fast helix-coil transition dynamics. This property, combined with its small size, makes WH5 a perfect candidate to carry out a systematic study of the kinetics of conformational rearrangements.[22–26]

Since we are interested in investigating the formation of a helical turn in the middle of the peptide chain, we classify the different WH5 conformations according to the arrangement of the three central alanines. Due to the fact that each one of these can adopt a coil or a helical configuration, this classification leads to 8 (2^3) different states. In this study we investigate the populations and the interconversion dynamics among all these states. For this purpose, we apply the recent development of well tempered metadynamics[27] proposed by Tiwary and Parrinello.[28] This approach, based on an infrequent deposition of the metadynamics bias, has already been successfully employed to obtain rates for processes such as binding/unbinding of the trypsin-benzamidine protein complex,[29] argon condensation,[30] and organics dissolution.[31]

While the dynamics of this process is fast enough to be studied by direct MD simulation, the use of the Tiwary and Parrinello approach alleviates the computational cost of our investigation. Furthermore this study provides yet another test of this metadynamics development in a case in which straight MD has been performed by us or by others.

Our study in agreement with previous results,[17,22,32–34] highlights the role that all-atom PFFs and especially WMs play in determining the dynamics of such systems.

II. MATERIAL AND METHODS

A. PFFs and WMs

To understand the influences of the PFFs and WMs on the helix-coil transition processes, we focus our attention on three amongst the best currently available PFFs:[12,14,19] Amber03w (A03), Charmm22* (C22), and Amber99SB* (A99). The first

a)ferruccio.palazzesi@phys.chem.ethz.ch

one has been introduced in 2010 by Best and Mittal,[35] in order to obtain in combination with TIP4P/2005,[36] a better description of the protein folding/unfolding processes. For C22, instead Piana *et al.*[17] re-parametrised the backbone torsional angles of the original Charmm22[37] to balance the helix-coil stability. Lastly, A99 is based on Amber99[38] and incorporates two different corrections: the SB parameters[39] to achieve a better control on the secondary structure stability and the Hummer and Best revisions[40] introduced to improve the reproduction of some helix/coil data. Recently it has been demonstrated that these PFFs in combination with specific WMs, generate fairly different conformational ensembles in the case of small and disordered peptides.[12,15,16] In combination with the aforementioned PFFs, we also consider three different explicit water models: TIP3P,[41] TIP4P/2005,[36] and the recently developed TIP4P-D.[20] In our study, we thus investigate a total of 9 different PFF/WM combinations. Since the results do not vary significantly as the TIP4P/2005 water model is changed in TIP4P-D, we report only the results obtained with the former one. The interested reader can find our results for the TIP4P-D model in the Tables S1, S2, and S4-S6 and Figures S3-S5, S7 and S9 of the supplementary material.

B. Representation of the conformational space

In order to distinguish between a coil (C) from a helix (H) conformation, we use the values of the backbone dihedral angles in combination with the Transition-Based Assignment (TBA).[42] This coarsened representation of the conformational space has already demonstrated its usefulness.[42,43] The application of this method allows to classify all 8 states of the WH5 system. We graphically represent these states as the vertices of a cube in Fig. 1. In such a scheme, the transitions between states are represented either as edges or diagonals. More details regarding the adopted representation of the configurational space are reported in Figure S1.

C. Enhanced MD simulations

To obtain the kinetics data on the folding and unfolding processes of a helix motif, we could have applied unbiased MD simulations, as routinely done in literature.[22,45–48] However to reduce the required computational time here we decided to use an enhanced sampling method such as metadynamics,[5,49] using an infrequent bias deposition. This is an approach based on the observation of Voter[50] and Grubmüller[51] that in a rare event scenario the unbiased transition rates can be computed from a bias calculation provided that zero bias is added to the transition state region. Tiwary and Parrinello[28] have shown how such a bias can be built in a metadynamics context. Very soon after, Salvalaglio *et al.*[52] have proposed a statistical test to verify whether the hypothesis under which the scheme of Tiway and Parrinello is valid is satisfied in the practice. This test is based on the consideration that in a rare event scenario, in which the residence time is much longer than the transition time, the distribution of escaping times has to be Poissonian. To further verify the validity of our calculations for one of our PFF/WM combination (A03+TIP4P/2005) we also run long unbiased MD simulations. See Table S3 for more details.

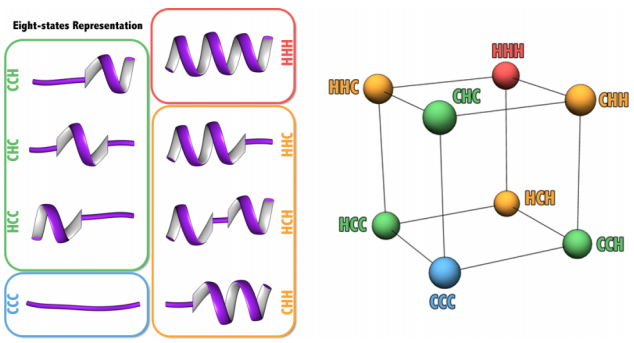

FIG. 1. In the left, we report a ribbon representation of all the possible configurations that the three central alanines of the WH5 peptide can adopt. In blue we have the all-coil state, in green the states in which one residue is in the helical conformation, in orange the states with two helical residues, while the all helix state is reported in red. On the right panel, we pictorially represent each of these states as vertexes of a cube. For the sake of clarity all the possible state-to-state transitions (black lines) are not shown.[44]

III. RESULTS AND DISCUSSION

A. Equilibrium state populations

In Fig. 2(a) we pictorially represent the equilibrium populations of the 8 states obtained, as described in Sec. II.

As can be seen from the figure, the conformational ensembles generated have comparable equilibrium populations, with low probabilities associated with the folded configurations. This result confirms the propensity of the system to assume in prevalence extended, unstructured conformations.[22]

A more quantitative representation of our results has been reported in Fig. 2(b), in which the equilibrium populations are aggregated according to the number of helical residues. From these data it is possible to measure differences between the system substate populations by varying the PFFs. In particular C22 favours the conformations with one helical residue. These differences are in absolute terms small. However, given the continuous strive at improving the all-atom empirical potentials we must consider them relevant.[14]

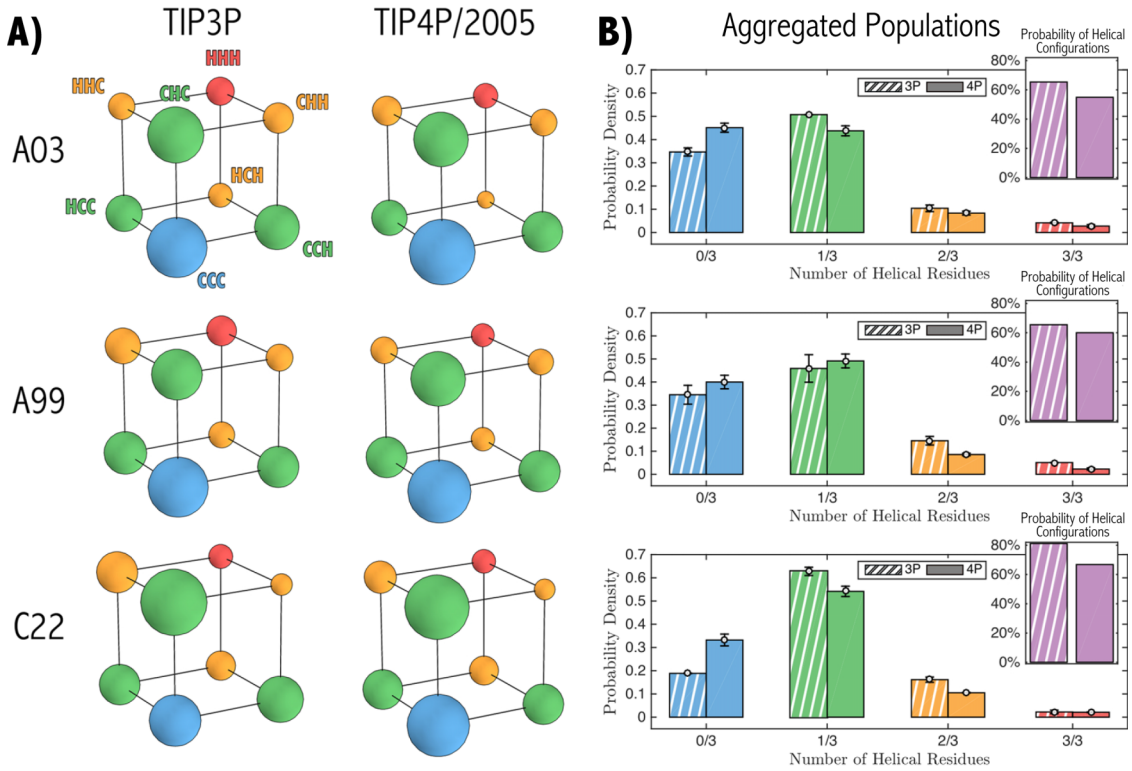

FIG. 2. (a) Illustrative comparison of the equilibrium state populations. The volume of the sphere is proportional to the state probability. (b) Distribution of the equilibrium populations aggregated according to the number of helical residues. In the inset, we report the probability of helical configurations. The bars with oblique white lines represent the TIP3P data, while the ones without lines represent the TIP4P/2005 data. The errors are calculated from block analysis.

Water models are known to exert some influence on the disordered state distributions,[20,32–34,58] and our results confirm this. From the data reported in Fig. 2, in fact, we observe that the TIP4P/2005 water model slightly favours less structured configurations. A consequence of this effect is a decrease of the probability of helical configurations for systems simulated with this water model. Such an effect can be consistently observed across all the PFFs considered.

As can be seen from Table S3 our results are in agreement to those obtained by De Sancho and Best[22] with a replica exchange technique, using one of the protein force-field and water model combinations investigated here.

B. Reaching the equilibrium

1. *Helix turn folding and unfolding rates*

We now move to analyse the helix nucleus formation and disruption rates, calculated for all the PFF/WM combinations using the procedure described in Sec. II. We report these data in Fig. 3. These estimates are in accordance with the folding time value calculated by De Sancho and Best[22] and with our own long unbiased MD simulations. See Table S3.

These findings show that the most important factor influencing the rate is the water model used and in particular TIP3P leads to a much faster kinetics.

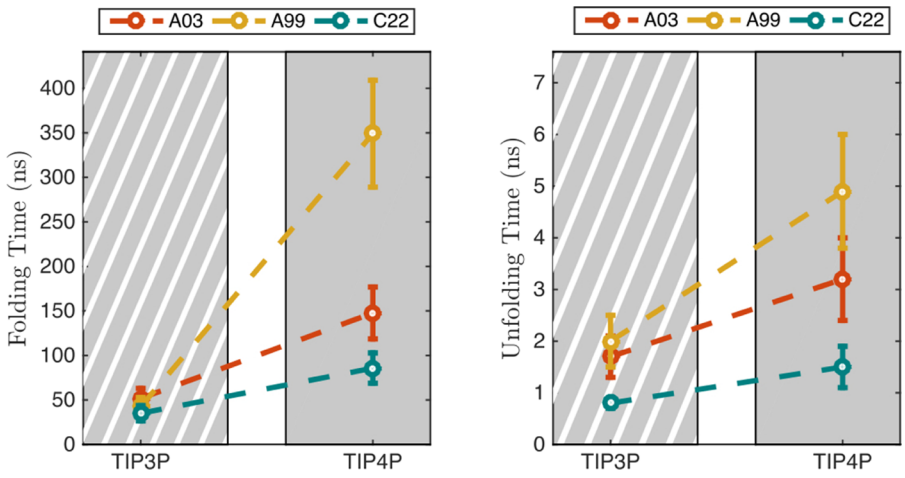

FIG. 3. On the left the helix turn folding times, while on the right the unfolding times. The error bars are obtained by performing 500 bootstrap tests.

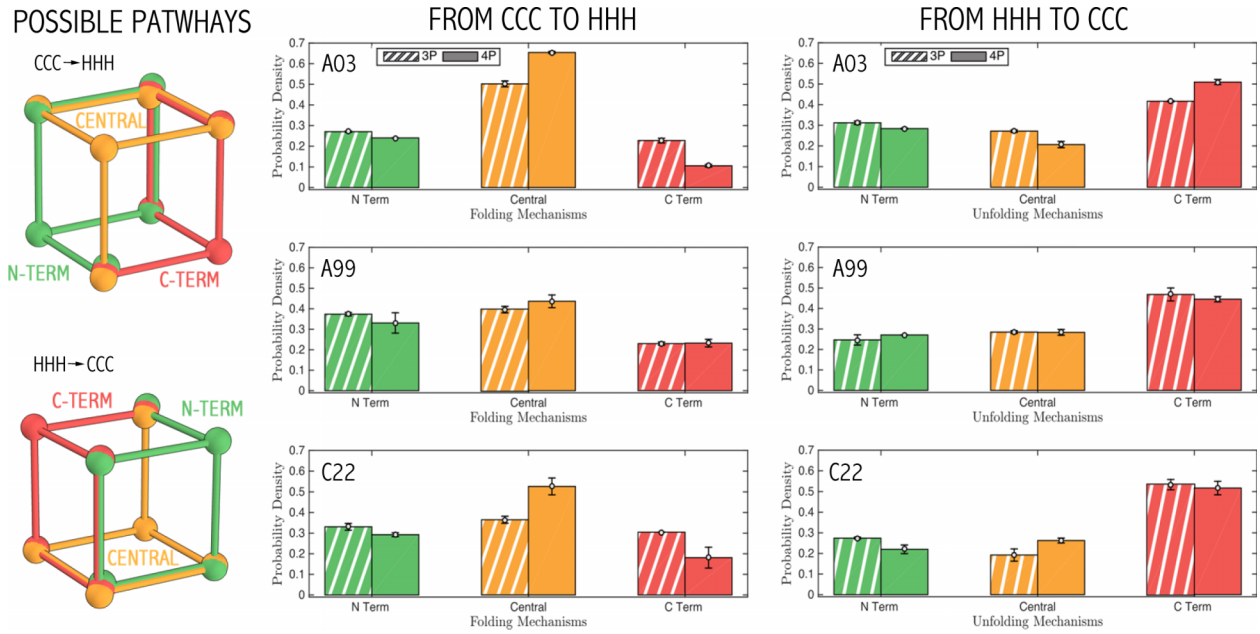

FIG. 4. Results of the analysis of the dominant helix turn folding/unfolding pathways. The bars with oblique white lines represent the TIP3P data, while the ones without lines represent the TIP4P/2005 data. The errors are calculated from block analysis.

In retrospect the rate dependence on the water model is not surprising.[32,59–62] In fact the TIP3P model is characterised by higher mobility, smaller surface tension, and smaller viscosity than any other TIP4P water models or not to mention real water.[20,32,36,63,64]

On the contrary, the protein force field used has a smaller role with C22 leading consistently to the slowest rates, while A99 to the fastest.

2. Helix turn folding and unfolding pathways

To calculate the dominant helix folding/unfolding pathways, we analyse all the state-to-state fluxes leaving the system free to evolve from a completely unfolded state (CCC) and from a completely folded one (HHH), and for this reason backward and forward fluxes differ (see Sec. II). For simplicity we have then grouped the dominant pathways into three classes, according to the intermediate states involved in these processes. Pathways that form or disrupt the helical conformation at the beginning of the WH5 peptide chain are assigned to the N-TERM category, while passages that involve the end residue are classified as C-TERM. Processes that change the conformations of the central alanine are instead considered as CENTRAL. A schematic representation of such a classification is reported in Fig. 4.

From such a figure we observe that while the WMs can strongly influence the time scale of the overall folding transition, the partition among different pathways is instead markedly influenced by the force-field, as previously observed.[17] The same holds for the reverse paths.

Previous computational work based on this peptide has also provided a detailed description of the helix folding/unfolding processes.[25,26] However, a different partition of the configuration space as well as the use of earlier generation of force-fields do not allow a direct comparison with our results.

IV. CONCLUSION

In this paper, using the recent developments[28,52] in the application of well tempered metadynamics,[27] we shed new light on helix turn folding/unfolding pathways of a small alanine-based peptide and the relative PFFs/WMs influences.

In order to understand the multiple effects of these empirical models and their different contributions, we systematically analyse different thermodynamic and kinetic properties, such as the equilibrium state distribution, the folding/unfolding rates, and the relative pathways. Our results suggest that WMs play a major role in the folding/unfolding process, affecting both the equilibrium and dynamics properties of the simulated system. Different explicit solvent models can in fact lead to different conformational ensemble distributions and different folding/unfolding times. Surprisingly, these observed differences do not influence the dominant pathways, probably due to the absence of specific solute-solvent interactions.

In this scenario, the choice of the WMs for an atomistic simulation emerges as a crucial choice. Moreover if we consider that flexible parts of the proteins are often able to play a direct role in many different biophysical processes;[65–68] a correct choice of the solvent is of great importance in obtaining a correct description of many different bio-related processes of interest.

Finally, the favourable comparison between the rates calculated with infrequent metadynamics and those obtained by straight MD provides another demonstration of the Tiwary and Parrinello scheme capabilities.[28]

SUPPLEMENTARY MATERIAL

See supplementary material for further details of the simulation protocol and the procedures adopted to calculate folding/unfolding rates, state-to-state fluxes, and dominant pathways from the MD trajectories.

ACKNOWLEDGMENTS

F.P. and M.P. acknowledge the EU-ERC Grant VARMET-670227 and SNF-MARVEL 51NF40 141828. A.B. acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-14-ACHN-0016. Calculations were carried out on Piz Dora and Piz Daint supercomputers at the Swiss National Supercomputing Center (CSCS) under project ID u1 and s638. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualisation, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

[1]V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature **390**, 196 (1997).
[2]J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Bioinf. **21**, 167 (1995).
[3]C.-J. Tsai, S. Kumar, B. Ma, and R. Nussinov, Protein Sci. **8**, 1181 (1999).
[4]B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. U. S. A. **97**, 8868 (2000).
[5]A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. **99**, 12562 (2002).
[6]J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Curr. Opin. Struct. Biol. **19**, 120 (2009).
[7]D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Willy, Science **330**, 341 (2010).
[8]A. E. García and J. N. Onuchic, Proc. Natl. Acad. Sci. U. S. A. **100**, 13898 (2003).
[9]X. Daura, W. F. van Gunsteren, and A. E. Mark, Proteins: Struct., Funct., Bioinf. **34**, 269 (1999).
[10]T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[11]K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011).
[12]F. Palazzesi, M. K. Prakash, M. Bonomi, and A. Barducci, J. Chem. Theory Comput. **11**, 2 (2014).
[13]R. B. Best, W. Zheng, and J. Mittal, J. Chem. Theory Comput. **10**, 5113 (2014).
[14]K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, J. Chem. Theory Comput. **8**, 1409 (2012).
[15]J. Henriques, C. Cragnell, and M. Skep, J. Chem. Theory Comput. **11**, 3420 (2015).
[16]J. A. Drake and B. M. Pettitt, J. Comput. Chem. **36**, 1275 (2015).
[17]S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Biophys. J. **100**, L47 (2011).
[18]P. S. Nerenberg and T. Head-Gordon, J. Chem. Theory Comput. **7**, 1220 (2011).
[19]K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, PloS One **7**, e32131 (2012).
[20]S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, J. Phys. Chem. B **119**, 5113 (2015).
[21]F. Martn-Garca, E. Papaleo, P. Gomez-Puertas, W. Boomsma, and K. Lindorff-Larsen, PLoS ONE **10**, e0121114 (2015).
[22]D. De Sancho and R. B. Best, J. Am. Chem. Soc. **133**, 6809 (2011).
[23]O. F. Mohammed, G. S. Jas, M. M. Lin, and A. H. Zewail, Angew. Chem., Int. Ed. **48**, 5628 (2009).
[24]M. M. Lin, O. F. Mohammed, G. S. Jas, and A. H. Zewail, Proc. Natl. Acad. Sci. U. S. A. **108**, 16622 (2011).
[25]G. S. Jas, W. A. Hegefeld, P. Májek, K. Kuczera, and R. Elber, J. Phys. Chem. B **116**, 6598 (2012).
[26]G. S. Jas and K. Kuczera, Mol. Simul. **38**, 682 (2012).
[27]A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).
[28]P. Tiwary and M. Parrinello, Phys. Rev. Lett. **111**, 230602 (2013).
[29]P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. **112**, E386 (2015).
[30]M. Salvalaglio, P. Tiwary, M. Mazzotti, and M. Parrinello, preprint arXiv: 1508.01642 (2015).
[31]J. Schneider and K. Reuter, J. Phys. Chem. Lett. **5**, 3859 (2014).
[32]P. Florová, P. Sklenovský, P. Banáš, and M. Otyepka, J. Chem. Theory Comput. **6**, 3569 (2010).
[33]D. S. Tomar, V. Weber, B. M. Pettitt, and D. Asthagiri, J. Phys. Chem. B **120**, 69 (2016).
[34]J. Henriques and M. Skepö, J. Chem. Theory Comput. **12**, 3407 (2016).
[35]R. B. Best and J. Mittal, J. Phys. Chem. B **114**, 14916 (2010).
[36]J. L. Abascal and C. Vega, J. Chem. Phys. **123**, 234505 (2005).
[37]A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[38]J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).
[39]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins: Struct., Funct., Bioinf. **65**, 712 (2006).
[40]R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009).
[41]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[42]N.-V. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).
[43]F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).
[44]E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, J. Comput. Chem. **25**, 1605 (2004).
[45]D. De Sancho, J. Mittal, and R. B. Best, J. Chem. Theory Comput. **9**, 1743 (2013).
[46]F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proc. Natl. Acad. Sci. U. S. A. **106**, 19011 (2009).
[47]N. Stanley, S. Esteban-Martín, and G. De Fabritiis, Nat. Commun. **5**, 5272 (2014).
[48]M. Shirts and V. S. Pande, Science **290**, 1903 (2000).
[49]A. Barducci, M. Bonomi, and M. Parrinello, WIREs: Comput. Mol. Sci. **1**, 826 (2011).
[50]A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
[51]H. Grubmüller, Phys. Rev. E **52**, 2893 (1995).
[52]M. Salvalaglio, P. Tiwary, and M. Parrinello, J. Chem. Theory Comput. **10**, 1420 (2014).
[53]R. Zwanzig, J. Stat. Phys. **30**, 255 (1983).
[54]J. D. Chodera and F. Noé, Curr. Opin. Struct. Biol. **25**, 135 (2014).
[55]N. P. Schafer, R. M. B. Hoffman, A. Burger, P. O. Craig, E. A. Komives, and P. G. Wolynes, PLoS ONE **7**, e50635 (2012).
[56]A. Berezhkovskii, G. Hummer, and A. Szabo, J. Chem. Phys. **130**, 205102 (2009).
[57]O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
[58]D. R. Nutt and J. C. Smith, J. Chem. Theory Comput. **3**, 1550 (2007).
[59]B. Zagrovic and V. Pande, J. Comput. Chem. **24**, 1432 (2003).
[60]Y. M. Rhee and V. S. Pande, J. Phys. Chem. B **112**, 6221 (2008).
[61]D. de Sancho, A. Sirur, and R. B. Best, Nat. Commun. **5**, 4307 (2014).
[62]W. Zheng, D. De Sancho, T. Hoppe, and R. B. Best, J. Am. Chem. Soc. **137**, 3283 (2015).
[63]M.-y. Shen and K. F. Freed, Biophys. J. **82**, 1791 (2002).
[64]M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 134508 (2005).
[65]M. M. Teeter, Annu. Rev. Biophys. Biophys. Chem. **20**, 577 (1991).
[66]M. S. Cheung, A. E. García, and J. N. Onuchic, Proc. Natl. Acad. Sci. U. S. A. **99**, 685 (2002).
[67]P. R. ten Wolde and D. Chandler, Proc. Natl. Acad. Sci. U. S. A. **99**, 6539 (2002).
[68]S. de Beer, N. P. Vermeulen, and C. Oostenbrink, Curr. Top. Med. Chem. **10**, 55 (2010).