

Let's get honest about sampling

David L. Mobley

Received: 8 November 2011 / Accepted: 15 November 2011 / Published online: 24 November 2011
© Springer Science+Business Media B.V. 2011

Abstract Molecular simulations see widespread and increasing use in computation and molecular design, especially within the area of molecular simulations applied to biomolecular binding and interactions, our focus here. However, force field accuracy remains a concern for many practitioners, and it is often not clear what level of accuracy is really needed for payoffs in a discovery setting. Here, I argue that despite limitations of today's force fields, current simulation tools and force fields now provide the potential for real benefits in a variety of applications. However, these same tools also provide irreproducible results which are often poorly interpreted. Continued progress in the field requires more honesty in assessment and care in evaluation of simulation results, especially with respect to convergence.

Keywords Sampling · Convergence · Force fields and molecular simulations · Binding affinity

Discussion

For molecular simulations to reliably predict, guide, and help explain experiment, these simulations require force fields of sufficient accuracy, adequate sampling of the relevant biomolecular motions (convergence) and a correct representation of the experimental conditions. Failures in any of these areas yield results which disagree with experiment. We may be tempted to blame disagreement with experiment on just one of these areas—force fields are perhaps the most common scapegoat, sometimes with good

reason [1–5]—but any or all of the three may be a weak point. And, in some sense, adequate sampling is the weakest link. Until sampling is adequate, equilibrium properties computed from a simulation remain biased by the system's starting state and no meaningful comparison with experiment is possible [6]. With an inadequate force field or a poor representation of the experimental conditions, results will disagree with experiment, but will be robust and improvement is relatively easy, but not so with inadequate sampling.

Many important biomolecular motions take place with characteristic timescales far longer than typical simulation timescales (even sidechain motions in the core of a protein can take hundreds of microseconds [7]), so one might expect that the literature would devote substantial attention to testing the adequacy of sampling in typical applications of molecular simulations to binding. However, this does not seem to be the case. Many errors get blamed on force field deficiencies, and perhaps more attention gets devoted to these, but at least in my own work on protein–ligand binding, the vast majority of the “accuracy” problems I have seen can be traced back to specific sampling problems, suggesting (at least in these systems) sampling may be a leading cause of error and thus that these are really problems of precision. Ligand binding modes are slow to change, presenting problems for binding mode prediction [6, 8–11]; protein conformational changes even at the single sidechain level can be slow, hurting the quality of computed binding free energies [12–14]; slow motion of waters into and out of binding sites can hurt convergence and thus apparent accuracy [3, 6, 15]; and unsampled protein conformational changes can also introduce errors [6]. Even ionic motions [16] and slow internal conformational changes in small molecules can pose problems [17–19]—on occasion, conformational energy barriers may be $14 k_B T$ even in small molecules [18–19]. These are all problems of

D. L. Mobley (✉)
Department of Chemistry, University of New Orleans,
2000 Lakeshore Drive, New Orleans, LA 70148, USA
e-mail: dmobley@gmail.com

timescales—typical simulations span the range of nanoseconds to (in heroic efforts) milliseconds [20], while important timescales for bimolecular rearrangements can be substantially longer—so these problems are perhaps not surprising.

Some recent efforts push the envelope in terms of simulation timescales, extending these out to milliseconds in some cases [20], with binding studies on the microsecond timescale [21, 22], which provides some grounds for enthusiasm. But even sidechain motions in the cores of proteins can be microsecond or slower events, while larger conformational changes and protein folding run even slower [7]. Perhaps as second-length simulations arrive on the scene in (hopefully) the next 25 years, we can be confident that sampling is adequate, but even then, we may begin seeing coupling between protein folding and ligand binding (such as in intrinsically disordered proteins) and sampling may *still* be a concern.

Given the potential for inadequate sampling, careful assessment of sampling is crucial for progress in the area. History demonstrates the importance of careful tests. Early work on binding prediction (using alchemical free energy calculations and other free energy techniques) saw some apparent high profile successes, resulting in considerable early enthusiasm which waned when it quickly became clear that the approach often yielded unreliable results that could be wildly wrong. This led to a lost decade (most of the 1990s) where these techniques saw relatively few applications outside of some of the key groups originating the techniques. Enthusiasm bounced back since 2001 or 2002. Obviously, this is less than ideal—steady (even if slow) progress is preferable.

To avoid similar cycles of enthusiasm, we must honestly assess sampling for adequacy. Despite the fact that many important biomolecular motions are almost guaranteed to be slower than typical simulation timescales, typical applications to biomolecular systems tend not to look very closely at this issue. In the best case scenario, a research group might begin multiple simulations from an identical set of starting structures to see whether they yield dramatically different results. This is better than no checking at all, but it is hardly a strenuous test of convergence, since these could all be starting in the same local minimum of the free energy landscape and remain trapped in that minimum on simulation timescales.

How should researchers look for convergence problems? Straightforward tests include starting from dramatically different starting structures (different crystal structures of the target receptor, or different homology models of the receptor, or substantially different structures generated from replica exchange type techniques [16], or several different potential ligand binding modes [6, 8, 23]), looking carefully for structural transitions, such as the number of sidechain torsional transitions in each residue around a

binding site in a receptor (and when this number is small but nonzero, it suggests inadequate sampling); and looking at cycle closure errors when computing free energies (such as in relative free energy calculations [24, 25]). More subtle convergence problems will certainly crop up as we push simulations to larger systems and longer timescales, and these may be harder to detect but of no less importance. In general, researchers should begin analysis with the assumption that typical simulation results remain un-converged, then construct simple tests to try and build up some confidence that results really are converged.

Force fields are undoubtedly important for accuracy, but inadequate sampling and convergence prevents meaningful comparison with experiment, so force fields can't even be accurately tested. In binding and free energy studies where we have obtained reasonable convergence, RMS errors relative to experiment have typically been in the 1–2 kcal/mol range [6, 8, 18, 25–27]. These levels of accuracy suffice for some benefits in discovery applications [25], depending on the workflow. Thus, a major bottleneck towards more widespread use of these techniques may not be force fields but rather convergence. With adequate sampling, we can quantitatively assess the accuracy of a particular force field, identify deficiencies, and improve it. Without adequate sampling, there is no such path forward.

Hence, simulations face a choice. We would like to plunge ahead and produce accurate and insightful results on a vast range of systems, and checking for convergence is hardly glamorous. But we must think more long-term. Where do we want to be in 25 years? Lack of short-term attention to convergence will yield simulation results which are irreproducible and unreliable, and follow-up work in the future will demonstrate this. If simulation is to gain trust and acceptance as a tool, convergence tests are essential. Otherwise, as we dash on to larger and larger systems, we will leave a trail of demonstrably poor convergence in our wake, fostering substantial backlash against simulations and moving them away from being a tool that sees widespread use.

Acknowledgements DLM acknowledges the Louisiana Board of Regents Research Competitiveness and Research Enhancement Subprograms as well as the Louisiana Optical Network Initiative (supported by the Louisiana Board of Regents Post-Katrina Support Fund Initiative grant LEQSF(2007-12)- ENH-PKSFI-PRS-01), and the National Science Foundation under NSF EPSCoR Cooperative Agreement No. EPS-1003897 with additional support from the Louisiana Board of Regents.

References

1. Merz KM Jr (2010) *J Chem Theory Comput* 6(5):1769
2. Faver JC, Benson ML, He X, Roberts BP, Wang B, Marshall MS, Kennedy MR, Sherrill CD, Merz KM Jr (2011) *J Chem Theory Comput* 7(3):790

3. Deng Y, Roux B (2008) *J Chem Phys* 128(11):115103
4. Fujitani H, Tanida Y, Matsuura A (2009) *Phys Rev E* 79(2):21914
5. Karney CFF, Ferrara JE, Brunner S (2005) *J Comput Chem* 26(3):243
6. Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK (2009) *J Mol Biol* 394(4):747
7. Schlick T (2010) *Molecular modeling and simulation: an interdisciplinary guide, interdisciplinary applied mathematics*, vol 21, 2nd edn. Springer, New York
8. Mobley DL, Graves AP, Chodera JD, McReynolds A, Shoichet BK, Dill KA (2007) *J Mol Biol* 371:1118
9. Jayachandran G, Shirts MR, Park S, Pande VS (2006) *J Chem Phys* 125(8):084901
10. Steinbrecher T, Case D, Labahn A (2006) *J Med Chem* 49(6):1837
11. Mobley DL, Dill KA (2009) *Structure* 17(4):489
12. Mobley DL, Chodera JD, Dill KA (2007) *J Chem Theory Comput* 3(4):1231
13. Gallicchio E, Levy RM (2011) *Curr Opin Struct Biol* 21(2):161
14. Gallicchio E, Lapelosa M, Levy RM (2010) *J Chem Theory Comput* 6(9):2961
15. Luccarelli J, Michel J, Tirado-Rives J, Jorgensen WL (2010) *J Chem Theory Comput* 6(12):3850
16. Michel J, Essex JW (2010) *J Comput Aided Mol Des* 24:638–658
17. Leitgeb M, Schröder C, Boresch S (2005) *J Chem Phys* 122(8):084109
18. Klimovich P, Mobley DL (2010) *J Comput Aided Mol Des* 24(4):307
19. Paluch AS, Mobley DL, Maginn EJ (2011) *J Chem Theory Comput* 7(9):2910
20. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) *Science* 334(6055):517
21. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H, Shaw DE (2011) *Proc Natl Acad Sci* 108(32):13118
22. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE (2011) *J Am Chem Soc* 133(24):9181
23. Mobley DL, Chodera JD, Dill KA (2006) *J Chem Phys* 125:084902
24. Dolenc J, Oostenbrink C, Koller J, van Gunsteren W (2005) *Nucleic Acids Res* 33(2):725
25. Shirts MR, Mobley D, Brown SP (2010) In: Merz KM, Ringe D, Reynolds CH (eds) *Drug design: structure- and ligand-based approaches*, Cambridge University Press
26. Mobley DL, Dumont É, Chodera JD, Dill K (2007) *J Phys Chem B* 111(9):2242
27. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) *J Chem Theory Comput* 5(2):350