

REVIEW

Open Access

High throughput molecular dynamics for drug discovery

Nathaniel Stanley¹ and Gianni De Fabritiis^{1,2*}

Abstract

Molecular dynamics simulations hold the promise to be an important tool for biological research and drug discovery. Historically, however, there were several obstacles for it to become a practical research tool. Limitations in computer hardware had previously made it difficult to simulate for long enough to see interesting biological processes. Recent improvements in hardware and algorithms have largely removed this issue, leaving data analysis as the main obstacle. Advances in Markov state modeling appear to be on the way to remove this obstacle. We outline these advances here and discuss numerous recent studies that demonstrate that molecular dynamics simulations will start to be an important tool for pharmaceutical research.

Keywords: High-throughput molecular dynamics; Markov state models; GPU; Drug discovery; Fragments

Review

Drug discovery is an iterative process that relies on various computational tools to help both lead experiments and understand data. High throughput virtual screening, docking, and hit development based on structure-activity relationships, are among many tools routinely used to identify or improve potential drug compounds (Jorgensen 2004; Sliwoski et al 2014). Still, most such methods rely on simplified assumptions that come with fundamental limitations. While such simplifications are needed at early stages, as development progresses and the chemical space begins to be narrowed down, methods with more accuracy should be employed (Rastelli et al 2009; Harvey et al 2009).

Molecular dynamics (MD) simulation is one such method. MD simulations combine Newtonian physics and an all-atom, flexible representation of proteins, water and other molecules to understand the dynamic interactions between them. They can provide important qualitative information, such as where and how a drug binds, but also quantitative information like the binding affinities and kinetics of such interactions (Buch et al 2011). This atomic-level description, combined with the

ability to compare it quantitatively with experiments has long made MD a very promising method.

Several substantial obstacles have traditionally prevented this from becoming a reality. Even the most basic biological events like side chain flipping or loop motions in a protein take hundreds of nanoseconds or longer (Zwier and Chong 2010). Therefore, many orders of magnitude in simulated time must be resolved in order to see even simple events in a single simulation. It is computationally costly to span so much time, and this has traditionally been what has kept MD from being practically useful for biological research and drug discovery. Enhanced sampling techniques were developed to speed up MD, but they require biasing along a reaction coordinate or prior knowledge of the system, which is many times unknown.

Further specialized supercomputers, such as the Anton supercomputer (Shaw et al 2007; Shaw et al 2014) or the MD-GRAPe (Ohmura et al 2014), have been developed that can run single simulations on very long timescales, up to a millisecond. A more practical way to approach this problem is to take advantage of recent hardware advances in GPU devices (Harvey et al 2009; Harvey and De Fabritiis 2012). A single GPU can now produce a microsecond simulation in a few days for a small system (~25,000 atoms). Running multiple parallel simulations on a small cluster of GPUs, one can reach millisecond aggregate sampling, a timescale needed to adequately

* Correspondence: gianni.defabritiis@upf.edu

¹Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, Barcelona 08003, Spain

²Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, Barcelona 08010, Spain

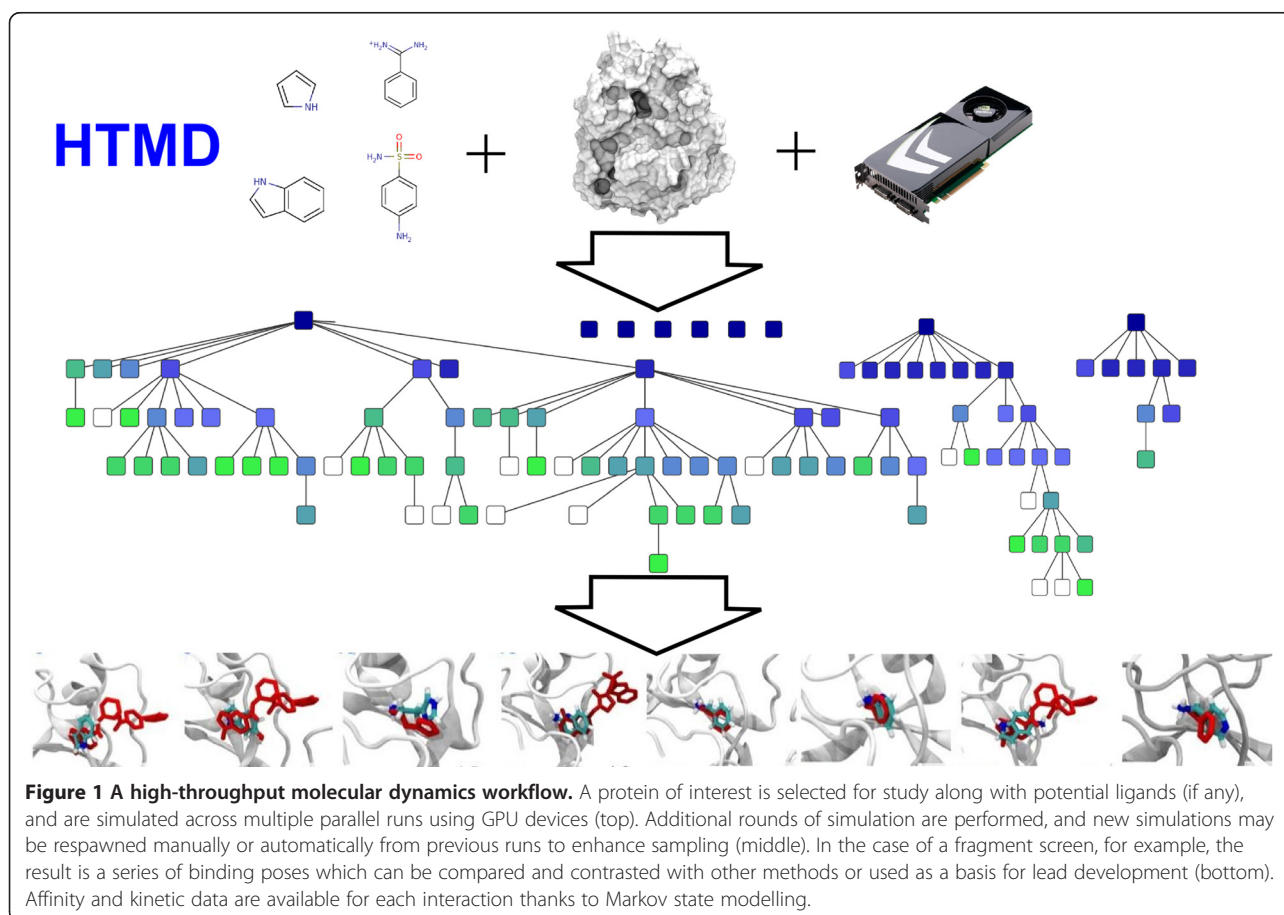
sample many biological processes, including binding of many small molecules (Buch et al 2010). We call this approach high-throughput molecular dynamics, or HTMD (Harvey and De Fabritiis 2012) (Figure 1). Accessibility to such hardware has never been easier to obtain thanks to commodity cloud services like Amazon AWS, so the barrier to entry to employing MD as a standard tool in the drug discovery process has been drastically lowered.

All this increased computational power results in a large amount of data, at which point analysis becomes a major concern. The copious and disjointed nature of the data produced by HTMD studies means that making sense of it is a significant challenge. Clustering methods to understand the data have trouble properly assigning weight and relevance to the data. Further, it is often not clear to newcomers that running multiple parallel simulations can allow one to investigate events that are much slower than the length of each individual simulation. This is indeed possible thanks to Markov state models (MSM), which allow one to take advantage of the statistical probability of events. The basics of MSMs have been covered at length, and we direct the reader to several publications for a more detailed look (Noé and Fischer 2008; Pande et al 2010; Prinz et al 2011). For the

remainder of this review, we focus mainly on our experience with these tools and proof-of-concept studies employing them.

One of the first studies to successfully use HTMD for ligand binding was Buch et al. (Buch et al 2011). Using just 50 μ s of simulation time, it was possible to reproduce the crystal binding pose, kinetics, and affinity of the binding of benzamidine to trypsin. However, while it was a critical proof-of-concept work, there were several aspects of the methods in that work which made it difficult to generalize to other systems. Others who worked on using MSMs for protein folding encountered similar successes and limitations, such as Bowman et al. found when studying Villin headpiece folding (Bowman et al 2009). While they could accurately approximate folding times, they found that RMSD based clustering was limited in part because structures that were close in RMSD may not interconvert rapidly, resulting in large heterogeneity inside clusters and hindering granularity of the MSM. Other clustering based on inter-atom contacts or distances, for example, has proved to be much more effective than spatial clustering.

Studies of folding and the motions of intrinsically disordered proteins provided additional difficulties. A



critical problem in this case was that clusters can be far away geometrically (different conformations), but kinetically very close. This would result in too many conformations which are kinetically close for the MSM to identify. A projection method known as time-sensitive Independent Component Analysis (tICA) (Schwantes and Pande 2013; Pérez-Hernández et al 2013) was therefore incorporated into the process before clustering in order to alleviate this problem. tICA projects the data along its slowest varying coordinates, which can then be fed to the clustering algorithms. This almost universally improves the accuracy of the results.

With all these improvements, several studies have been able to show that HTMD can now be incorporated into a drug discovery pipeline. In an as yet unpublished work, we have recently measured affinity and kinetics for 42-fragment screen targeting Factor Xa, identifying crystal structures as well as secondary happy poses which are key to properly interpret experimental data on fragments. Expanding these methods to a full fragment library, typically on the order of 700 fragments, would be a simple question of cost and therefore time, considering the ever increase power of computing resources. Whether that is practical and cheaper than current best practices like X-ray and NMR spectroscopy remains to be answered.

Studies of binding in membrane proteins are also possible, even in difficult cases where the ligand is a lipid itself. In two studies by our group, we were able to show the binding pathway of lipid ligands binding to target proteins. In one work, we simulated the binding of the lipid anandamide to the enzyme FAAH (Dainese et al 2014). We also demonstrated how cholesterol interacts with and modulates the enzyme. In another, still unpublished work, we demonstrated the mechanism of binding of a lipid inhibitor, ML056, to sphingosine-1-phosphate receptor 1 (S1PR1). In that work, we were able to reproduce the crystallographic binding pose of the inhibitor, as well as characterize several important conformational changes along the pathway to binding. These studies were computationally intensive, requiring 250 μ s and 800 μ s of simulation respectively, and would be effectively impossible without a HTMD paradigm using standard hardware.

Drug design is more than just molecular recognition and binding. Often some fundamental activity of the protein remains to be understood before a drug discovery initiative can even start. MD has also been used to demonstrate that a postulated mechanism for HIV protease to cleave itself out of a long protein chain was indeed correct (Sadiq et al 2012). Novel behavior in the KID disordered protein (Stanley et al 2014) was also unveiled by a massive use of HTMD simulations. Accurately assessing the kinetic properties of both these

systems was important and required 335 μ s and 1.7 ms of simulation, respectively. And as a further example of how simulations can be important for therapeutic discovery, Shan et al. have used them to explain how mutations to EGFR lead to cancers (Shan et al 2012; Shan et al 2013).

Conclusion

Computational tools are incorporated into the drug discovery pipeline because they speed up or corroborate experimental tests along the iterative cycle towards a drug. Molecular dynamics simulations are now accurate and fast enough that they can be used to help guide design choices of potential drugs. HTMD studies allow the full binding process of a compound to be sampled, giving important details on transient interactions, kinetics, affinity, and final resting pose. It can be used to test and rank an array of fragment compounds or analyze the binding of a lead compound. Further, they can be used to understand basic biophysical behavior that may be important before drug design even begins. As the raw power of individual GPUs increases and cloud services become ever cheaper and more routinely accessible, such studies will become even more commonplace.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NS and GDF wrote the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We thank the volunteers of GPUGRID.net for contributing compute time to some of the works mentioned in this review.

Received: 23 January 2015 Accepted: 4 February 2015

Published online: 13 February 2015

References

- Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *J Chem Phys* 131:124101, doi:10.1063/1.3216567
- Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. *J Chem Inf Model* 50:397–403, doi:10.1021/ci900455r
- Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A* 108:10184–10189, doi:10.1073/pnas.1103547108
- Dainese E, De Fabritiis G, Sabatucci A, Oddi S, Angelucci CB, Di Pancrazio C, Giorgino T, Stanley N, Del Carlo M, Cravatt BF, Maccarrone M (2014) Membrane lipids are key modulators of the endocannabinoid-hydrolase FAAH. *Biochem J* 457:463–472, doi:10.1042/BJ20130960
- Harvey MJ, De Fabritiis G (2012) High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug Discov Today* 17:1059–1062, doi:10.1016/j.drudis.2012.03.017
- Harvey MJ, Giupponi G, Fabritiis GD (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* 5:1632–1639, doi:10.1021/ct9000685
- Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303:1813–1818, doi:10.1126/science.1096361
- Noé F, Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* 18:154–162, doi:10.1016/j.sbi.2008.01.008

- Ohmura I, Morimoto G, Ohno Y, Hasegawa A, Taiji M (2014) MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. *Philos Trans R Soc Math Phys Eng Sci* 372:20130387, doi:10.1098/rsta.2013.0387
- Pande VS, Beauchamp K, Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52(1):99–105, doi: 10.1016/j.jymeth.2010.06.002
- Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139:015102–015113, doi:10.1063/1.4811489
- Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F (2011) Markov models of molecular kinetics: generation and validation. *J Chem Phys* 134:174105–174123, doi:10.1063/1.3565032
- Rastelli G, Degliesposti G, Del Rio A, Sgobba M (2009) Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chem Biol Drug Des* 73:283–286, doi:10.1111/j.1747-0285.2009.00780.x
- Sadiq SK, Noé F, Fabritiis GD (2012) Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc Natl Acad Sci U S A* 109(50):20449–20454, doi: 10.1073/pnas.1210983109
- Schwantes CR, Pande VS (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J Chem Theory Comput* 9:2000–2009, doi:10.1021/ct300878a
- Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE (2012) Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* 149:860–870, doi:10.1016/j.cell.2012.02.063
- Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE (2013) Transitions to catalytically inactive conformations in EGFR kinase. *Proc Natl Acad Sci U S A* 110:7270–7275, doi:10.1073/pnas.1220843110
- Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC (2007) Anton, a special-purpose machine for molecular dynamics simulation. In: *Proc. 34th annu. int. symp. comput. archit.* ACM, New York, NY, USA, pp 1–12
- Shaw DE, Grossman JP, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH, Forte A, Gagliardo J, Gill G, Greskamp B, Ho CR, Ierardi DJ, Iserovich L, Kuskin JS, Larson RH, Layman T, Lee L-S, Lerer AK, Li C, Killebrew D, Mackenzie KM, Mok SY-H, Moraes MA, Mueller R, Nociolo LJ, Peticolas JL et al (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* IEEE Press, Piscataway, NJ, USA, pp 41–53
- Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational methods in drug discovery. *Pharmacol Rev* 66:334–395, doi:10.1124/pr.112.007336
- Stanley N, Esteban-Martín S, De Fabritiis G (2014) Kinetic modulation of a disordered protein domain by phosphorylation. *Nat Commun* 5:5272, doi: 10.1038/ncomms6272
- Zwier MC, Chong LT (2010) Reaching biological timescales with all-atom molecular dynamics simulations. *Curr Opin Pharmacol* 10:745–752, doi:10.1016/j.coph.2010.09.008

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
