Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2017 February 01.

Published in final edited form as:

Curr Opin Struct Biol. 2016 February ; 36: 25–31. doi:10.1016/j.sbi.2015.12.002.

Advances in free-energy-based simulations of protein folding and ligand binding

Alberto Perez[1], **Joseph A. Morrone**[1], **Carlos Simmerling**[1,2], and **Ken A. Dill**[1,2,3]

[1]Laufer Center for Physical and Quantitative Biology

[2]Department of Chemistry, Stony Brook University

[3]Department of Physics and Astronomy, Stony Brook University

Abstract

Free-energy-based simulations are increasingly providing the narratives about the structures, dynamics and biological mechanisms that constitute the fabric of protein science. Here, we review two recent successes. It is becoming practical: (i) to fold small proteins with free-energy methods without knowing substructures, and (ii) to compute ligand-protein binding affinities, not just their binding poses. Over the past 40 years, the timescales that can be simulated by atomistic MD are doubling every 1.3 years – which is faster than Moore's law. Thus, these advances are not simply due to the availability of faster computers. Force fields, solvation models and simulation methodology have kept pace with computing advancements, and are now quite good. At the tip of the spear recently are GPU-based computing, improved fast-solvation methods, continued advances in force fields, and conformational sampling methods that harness external information.

Introduction

Increasingly, our understanding of the properties and actions of proteins depends upon physics-based molecular simulations. It is of interest to model the folding of proteins into their stable structures, the binding affinities and selectivities of ligand-protein and protein-protein assemblies, as well as the solubilities and partitioning of biomolecules. The principled way to predict either static properties, or nanosecond-by-nanosecond and Angstrom-by-Angstrom detailed narratives of these actions is to utilize techniques that sample from the free energy surface and reflect thermal populations.

Different approaches can be taken to model biomolecules. Much can be inferred about protein structures by purely comparative modeling using the known structures in the PDB. Often, however equilibrium and kinetic information is desired. These can be inferred within a single framework using force fields and solvent models combined with sampling methods such as molecular dynamics (MD) or Monte Carlo (MC). While direct application of such methods can, in principle, properly sample populations and identify stable states, MD or MC

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

simulations alone are usually not capable of traversing barriers and sampling rare events sufficiently to quantify the free energy differences among states. Molecular simulations can be coupled with specialized techniques for enhancing sampling or extracting information from multiple equilibrium states for this purpose, and these are often referred to as “free-energy calculations.” Here we take the view that all methods that sample the free energy surface are “free-energy-based simulation,” ranging from “brute force” MD to methods that better facilitate barrier crossings such as replica exchange MD, to techniques such as free-energy perturbation or umbrella sampling.

The obstacles to free-energy modeling have been its high computational cost and some physical inaccuracies in the energetics. However, our current opinion is that free energy methods have become powerful both rapidly and recently. First, Figure 1 shows that advances in computers, force fields, and methodology over the past 40 years have led to faster-than-Moore's-Law increases in the timescales that are now accessible to simulation. Second, here we review progress on two key problems – predicting protein native structures and ligand-binding affinities. These are but two examples and there are many others we do not have the space to cover here.

Atomistic simulations are now folding small proteins and predicting their native structures

There are recent successes in computing the native structures of small proteins by physics-based methods, for diverse folds, without direct inputs from structural databases or the need for structural alignments. DE Shaw Research (DESRES) showed that a single atomistic force field can give folding trajectories for 12 small proteins over long-time simulated trajectories in explicit water on Anton, a special purpose supercomputer [6] [7]. This was an important milestone in proving the power and transferability of a current force field, and in computation of the folding pathways, some of which have been confirmed experimentally [8] [9]. However, their goal was not so much to obtain accurate populations, as it was to sample multiple folding events, so their simulations were performed near the melting temperature.

A complementary study by Nguyen, et al. was aimed at detecting native structures starting from fully extended conformations. Nguyen et al. used implicit-solvent and required only lab-sized computer clusters [10]. They attributed their successes to the use of GPU-based computing, and to good implicit solvent [11] and force field models [12]. Native structures were found reliably for proteins up to 50 residues, but predictions were not as consistent for longer proteins (up to 92 residues in the study). The main challenge with larger proteins was shown to be limitations of the sampling, not the force field. And, importantly, their implicit-solvent performed as well in this test as previous, more expensive, explicit modeling, implying the adequacy of fast solvation models for some protein modeling previously thought to require more computationally expensive models.

A huge challenge for finding native structures by atomistic MD is that folding times increase sharply with protein size. Experiments show that the folding time increases exponentially with the square root of the number amino acids [13]. Force field limitations and sampling

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inefficiency make the problem more daunting for simulation. This challenge motivates a need for free-energy methods that can search efficiently to find basins of important states, and that can sample well the populations within those basins.

Recently, an approach called MELD (Modeling Employing Limited Data), has been developed to find and sample important states efficiently by “melding” structural or heuristic information into MD simulations [14,15]. MELD is able to use combinatorically vague and generic instructives such as: “make a hydrophobic core” or “make secondary structures that are consistent with those provided by webservers” or “make a compact structure.” MELD accelerates conformational searching substantially, while at the same time preserving its critically important ability to give free energies. For example, MELD finds and samples well native structures (better than 4 Å RMSD) for 15 out of 20 small proteins, up to 92-mers, starting from fully extended states [15]. The speed advantage of MELD over brute-force MD is shown in Figure 2A.

Figure 2B shows the implication for computational structure prediction going forward: Even with future Moore's-law-like advances, the severity of the exponential search problem means that pure MD will not be folding proteins bigger than 140-mer proteins for another 25 years. But, when MD is combined with external information, as is done in MELD with generic instructives, Fig 2B projects that free energy methods will give native structures of those sizes within just 4-5 years. This protein size covers a large fraction of single domain proteins, including many biologically relevant ones like ribonuclease (134 residues), lysozyme (129), calmodulin (148) or myoglobin (154).

Free-energy methods are predicting the binding affinities of small ligands to simple proteins

Computational drug discovery is also poised to benefit from advances in free-energy methods. A traditional method for computational drug discovery has been DOCK and related algorithms [16-18]. Docking methods are fast and are often able to find correct binding site and ligand poses, but they rarely give accurate binding affinities.

Although more expensive than docking, physics-based methods have long promised to predict more accurate binding affinities, because of their better potentials and more complete conformational sampling and solvation. In a recent advance, DESRES [19] [20] and others [21] have run long MD simulations, and observed the ligand seeking and finding its binding site on the protein. Such work highlights the ability of atomistic simulations to identify stable states given no prior knowledge of the binding site or pose.

Even so, it is not yet possible to sample enough unbinding events to determine rates or affinities by direct MD. Specialized free energy methods are used to enhance or accelerate sampling, sometimes using “alchemical” pathways that transform one species to another. Techniques such as free energy perturbation [22], umbrella sampling [23], and the multi-state Bennett acceptance ratio [24] have been used to determine free energy differences. The goal is usually to compute one of two quantities: (i) *The relative* binding free energy (RBFE), which is the affinity of a ligand, given experimental knowledge of a similar

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reference ligand, or (ii) *The absolute* binding free energy (ABFE), which is the free energy of binding a ligand, without knowledge of any experimental reference compound. Although the latter requires less prior added knowledge, it is computationally more expensive to evaluate. Early studies have shown that free energy methods can be used to compute solvation free energies [25] and relative ligand binding free energies [26-29]. More recently, absolute binding free energies have been computed by methods of mean force or alchemical transformation [30-33].

Recently, systematic improvements in force fields and free energy techniques have begun to pay off and more diverse sets of systems can be treated accurately. Fig. 3A shows RBFE simulations of 8 proteins bound to nearly 200 different ligands, where the RMS errors are less than 2 kcal/mol [34]. Fig. 3B shows ABFE computations of Cyt C Peroxidase bound to 19 different ligands [35]. These, too, give RMS errors within about 2 kcal/mol. There are also recent successes using the sampling methods of metadynamics [36] and accelerated MD [37]. An unexpectedly welcome finding [38] is that implicit solvation may be sufficient for computing binding free energies – as we noted above that it may be for protein structure prediction – promising significant speed improvements going forward.

Another important application of simulation techniques is to compute *ligand-binding off rates.* A ligand's off rate (the inverse of its residence time on the protein) is sometimes better correlated with drug efficiency than binding affinities are [40,41]. For *Mycobacterium tuberculosis* enoyl-ACP, MD simulations and free energy calculations have identified the transition state to slow-onset inhibition. Such calculations now enable the rational modification of inhibitors or proteins based on their desired residence time [42,43]. Tiwary, et al. have computed the rate of unbinding of a trypsin-benzamidine complex using “infrequent metadynamics” techniques [44]. Theory and simulation have also illuminated the role of solvation and desolvation in ligand binding kinetics [45,46].

Challenges remain in calculating binding affinities. For example, when ligands cause induced fit in the protein, small changes can sometimes be sampled [34,47], but larger ones usually cannot. In addition, despite some encouraging results (Fig. 2b), charged ligands and charged binding sites remain a major challenge in general [35].

Force fields continue to improve, but still have limitations

Force fields and solvent models are advancing by bootstrapping. As they get better and faster, they are tested on larger databases and on systems that are bigger, more challenging, and more biological, which, in turn, leads to further improvements. Force fields advances are driven by increased computer power, particularly GPUs; the use of training against higher levels of QM; the inclusion of more conformational diversity in training sets; the use of bigger peptide-sized molecules, rather than just small organics; and the use of better databases for testing, such as of NMR scalar couplings [48] and amino-acid helical propensities [49,50]. In earlier protein-folding modeling, force fields were systematically unbalanced between different types of secondary structures. Newer models have much better balance, as shown by recent studies where a variety of protein structures can be folded with a single force field [51,52] [6,10,15]. There remain challenges in capturing the secondary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

structure propensities of different amino acids and accurately modeling unfolded or disordered proteins [53,54] – a research area receiving recent attention [55,56,57]. And, it remains challenging to model disordered proteins, because of the extensive sampling needed.

Although most current MD force fields are purely physics-based, some force fields, such as CHARMM[58], also include empirical fitting against experimental data in the final stages of their development. This kind of adjustment can help offset weaknesses in the force field functional form and improve the accuracy of the physics. Although the Amber force fields have not typically included such empirical adjustments, several groups have found that they improve agreement with experiments [59,60]. The need for empirical adjustments may indicate that force fields are reaching the limits of their simple functional forms, as compared to more computationally demanding models that include explicit polarizability [61,62]. For example, it was shown that polarizable models can give a better description of the cooperativity of α-helix formation [63]. These subtle effects can be difficult or impossible to include in simpler models, even with empirical modifications. As an alternative to empirical adjustment, other studies have suggested more fundamental revisions of additive force fields, such as new charge models [64], or using aqueous solvent in the QM fitting data [65]. However, these developments remain relatively untested.

Our Current Opinion: the power of free-energy simulations continues to grow

Free-energy computer simulations are giving increasing insights into proteins through the modeling of the populations, stable states, dynamics, motions, binding and partitioning that provide the language for how we understand protein machines. Over the past 40 years, atomistic simulations have progressed faster than Moore's Law in the timescales they can simulate, and they have also advanced in accuracy and narrative power. Some of this progress is attributable to advances in computers – IBM Blue Gene, the Anton supercomputer [7], and Folding@Home and other distributed grid systems [66], and more recently to GPU technology. However, much credit is also due to improved MD force fields, methods and solvation models [67,68], and to new types of sampling that harness external information within atomistic simulations. MD is now predicting small protein structures, without the assistance of structural databases, as well as the affinities of ligands binding to simple proteins. A remarkable bit of good fortune is that implicit-solvent models, which are computationally inexpensive, appear to be sufficient in some of these applications.

The systematic progress in computational biophysics is beginning to pay off in increasingly bigger ways for structural biology. As the time and length scales that can be simulated continue to grow, so does our ability to tackle more challenging problems. The current frontier includes modeling protein-protein and protein-DNA binding affinities, the behavior of ATPases [69-71], aggregation and amyloid formation [72], and the cytoplasm of a bacterium cell [73].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgements

This work was supported by NIH grants GM090205 and GM107104. The authors appreciate support from the Laufer Center.

References

1. Wikipedia. Transistor count. https://en.wikipedia.org/wiki/Transistor_count
2. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977; 267:585–590. [PubMed: 301613]
3. van Gunsteren WF, Karplus M. Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. Biochemistry. 1982; 21:2259–2274. [PubMed: 6178423]
4. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science. 1998; 282:740–744. [PubMed: 9784131]
5. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. 2008; 94:L75–7. [PubMed: 18339748]
6. Piana S, Lindorff-Larsen K, Piana S, Dror RO, Shaw DE, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011; 334:517–520. [PubMed: 22034434]
7. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. Anton, a special-purpose machine for molecular dynamics simulation. ACM SIGARCH Computer Architecture News. 2007; 35:1–12.
8. Prigozhin MB, Gruebele M. Microsecond folding experiments and simulations: a match is made. Phys Chem Chem Phys. 2013; 15:3372–3388. [PubMed: 23361200]
9*. Prigozhin MB, Chao S-H, Sukenik S, Pogorelov TV, Gruebele M. Mapping fast protein folding with multiple-site fluorescent probes. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:7966–7971. [PubMed: 26080403] [This article compares fast temperature jump fluorescent experiments with previously publihsed long molecular dynamics folding simulations concluding that the two can be compared at the same timescale.]
10. Nguyen H, Maier J, Huang H, Perrone V, Simmerling C. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. J. Am. Chem. Soc. 2014; 136:13959–13962. [PubMed: 25255057]
11. Nguyen H, Roe DR, Simmerling C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. J Chem Theory Comput. 2013; 9:2020–2034. [PubMed: 25788871]
12. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015; 11:3696–3713. [PubMed: 26574453]
13. Naganathan AN, Muñoz V. Scaling of folding times with protein size. J. Am. Chem. Soc. 2005; 127:480–481. [PubMed: 15643845]
14. MacCallum JL, Perez A, Dill K. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:6985–6990. [PubMed: 26038552]
15. Perez A, MacCallum JL, Dill K. Accelerating molecular simulations of proteins using Bayesian inference on weak information. [Internet]. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:11846–11851. [PubMed: 26351667]
16. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, Rizzo RC. Development and validation of a modular, extensible docking program: DOCK 5. J Comput Aided Mol Des. 2006; 20:601–619. [PubMed: 17149653]
17. Friesner, Richard A.; Banks, Jay L.; Murphy, Robert B.; Halgren, Thomas A.; Klicic, Jasna J.; Mainz, Daniel T.; Repasky, Matthew P.; Knoll, Eric H.; Shelley, Mee; Perry, Jason K., et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J. Med. Chem. 2004; 47:1739–1749. [PubMed: 15027865]
18. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. Proteins. 2003; 52:609–623. [PubMed: 12910460]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

19. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How Does a Drug Molecule Find Its Target Binding Site? J. Am. Chem. Soc. 2011; 133:9181–9183. [PubMed: 21545110]
20. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H, Shaw DE. Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:13118–13123. [PubMed: 21778406]
21. Giorgino T, Buch I, De Fabritiis G. Visualizing the Induced Binding of SH2-Phosphopeptide. J Chem Theory Comput. 2012; 8:1171–1175. [PubMed: 26596736]
22. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. 1954; 22:1420–1426.
23. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics. 1977; 23:187–199.
24. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 2008; 129:124105. [PubMed: 19045004]
25. Jorgensen WL, Ravimohan C. Monte Carlo simulation of differences in free energies of hydration. J. Chem. Phys. 1985; 83:3050.
26. Tembre BL, Mc Cammon JA. Ligand-receptor interactions. Computers & Chemistry. 1984; 8:281–283.
27. Wong CF, McCammon JA. Dynamics and Design of Enzymes and Inhibitors. J. Am. Chem. Soc. 1986; 108:3830–3832.
28. Bash PA, Singh UC, Brown FK, Langridge R, Kollman PA. Calculation of the Relative Change in Binding Free-Energy of a Protein-Inhibitor Complex. Science. 1987; 235:574–576. [PubMed: 3810157]
29. Bash PA, Singh UC, Langridge R, Kollman PA. Free energy calculations by computer simulation. Science. 1987; 236:564–568. [PubMed: 3576184]
30. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. J. Mol. Biol. 2007; 371:1118–1134. [PubMed: 17599350]
31. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. Proc. Natl. Acad. Sci. U.S.A. 2005; 102:6825–6830. [PubMed: 15867154]
32. Deng Y, Roux B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. J Chem Theory Comput. 2006; 2:1255–1273. [PubMed: 26626834]
33. Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get.”. Structure/Folding and Design. 2009; 17:489–498. [PubMed: 19368882]
34**. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J. Am. Chem. Soc. 2015; 137:2695–2703. [PubMed: 25625324] [This paper presents a large set of accurate, state-of-the-art relative binding free energy calculations.]
35. Rocklin GJ, Boyce SE, Fischer M, Fish I, Mobley DL, Shoichet BK, Dill K. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. J. Mol. Biol. 2013; 425:4569–4583. [PubMed: 23896298]
36. Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:6358–6363. [PubMed: 23553839]
37. Kappel K, Miao Y. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. Quarterly Reviews of Biophysics. 2015 doi:10.1017/S0033583515000153.
38*. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. J Comput Aided Mol Des. 2013; 27:989–1007. [PubMed: 24297454] [This paper presents a study of ligand binding to the T4 lysozyme L99A model system using efficient implicit solvent models that facilitate the sampling of multiple binding sites and binding poses.]
39.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Panel A adapted with permission from Ref. [34]. Copyright 2015 American Chemical Society. Panel B adapted with permission from Ref. [35]. Copyright 2013, with permission from Elsevier.

40. Lu H, Tonge PJ. Drug-target residence time: critical information for lead optimization. Current Opinion in Chemical Biology. 2010; 14:467–474. [PubMed: 20663707]
41. Copeland RA, Pompliano DL, Meek TD. Drug–target residence time and its implications for lead optimization. Nat Rev Drug Discov. 2006; 5:730–739. [PubMed: 16888652]
42. Li H-J, Lai C-T, Pan P, Yu W, Liu N, Bommineni GR, Garcia-Diaz M, Simmerling C, Tonge PJ. A Structural and Energetic Model for the Slow-Onset Inhibition of the Mycobacterium tuberculosis Enoyl-ACP Reductase InhA. ACS Chem. Biol. 2014; 9:986–993. [PubMed: 24527857]
43. Lai C-T, Li H-J, Yu W, Shah S, Bommineni GR, Perrone V, Garcia-Diaz M, Tonge PJ, Simmerling C. Rational Modulation of the Induced-Fit Conformational Change for Slow-Onset Inhibition in Mycobacterium tuberculosis InhA. Biochemistry. 2015; 54:4683–4691. [PubMed: 26147157]
44*. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:E386–91. [PubMed: 25605901] [This paper presents a calculation of the protein-ligand unbinding (off) rate using novel infrequent metadyanmics techniques.]
45. Setny P, Baron R, Michael Kekenes-Huskey P, McCammon JA, Dzubiella J. Solvent fluctuations in hydrophobic cavity–ligand binding kinetics. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:1197–1202. [PubMed: 23297241]
46. Mondal J, Morrone JA, Berne BJ. How hydrophobic drying forces impact the kinetics of molecular recognition. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:13277–13282. [PubMed: 23901110]
47. Mobley DL, Chodera JD, Dill K. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. J Chem Theory Comput. 2007; 3:1231–1235. [PubMed: 18843379]
48. Graf J, Nguyen PH, Stock G, Schwalbe H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. J. Am. Chem. Soc. 2007; 129:1179–1189. [PubMed: 17263399]
49. Best RB, de Sancho D, Mittal J. Residue-specific α-helix propensities from molecular simulation. Biophys. J. 2012; 102:1462–1467. [PubMed: 22455930]
50. Moreau RJ, Schubert CR, Nasr KA, Torok M, Miller JS, Kennedy RJ, Kemp DS. Context-independent, temperature-dependent helical propensities for amino acid residues. J. Am. Chem. Soc. 2009; 131:13107–13116. [PubMed: 19702302]
51. Best RB, Mittal J. Balance between a and beta Structures in Ab Initio Protein Folding. J Phys Chem B. 2010; 114:8790–8798. [PubMed: 20536262]
52. Mittal J, Best RB. Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. Biophys. J. 2010; 99:L26–L28. [PubMed: 20682244]
53*. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr. Opin. Struct. Biol. 2014; 24:98–105. [PubMed: 24463371] [This review addresses the success and failures of current force fields in long molecular dynamics simulations to agree with experimental evidence.]
54. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic Validation of Protein Force Fields against Experimental Data. PLoS One. 2012; 7:e32131. [PubMed: 22384157]
55. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015; 119:5113–5123. [PubMed: 25764013]
56. Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014; 10:5113–5124. [PubMed: 25400522]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

57. Mercadante D, Milles S, Fuertes G, Svergun DI, Lemke EA, Gräter F. Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. J Phys Chem B. 2015; 119:7975–7984. [PubMed: 26030189]

58. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. 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 Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]

59. Wang W, Ye W, Jiang C, Luo R, Chen H-F. New force field on modeling intrinsically disordered proteins. Chem Biol Drug Des. 2014; 84:253–269. [PubMed: 24589355]

60. Perez A, MacCallum JL, Brini E, Simmerling C, Dill K. Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations. J Chem Theory Comput. 2015; 11:4770–4779. [PubMed: 26574266]

61*. Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder JW, Ren P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. J Chem Theory Comput. 2013; 9:4046–4063. [PubMed: 24163642] [AMOEBA presents one of the initiatives to go beyond current single point charge force fields to introduce polarizability.]

62*. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B, MacKerell AD. Force Field for Peptides and Proteins based on the Classical Drude Oscillator. J Chem Theory Comput. 2013; 9:5430–5449. [PubMed: 24459460] [This paper represents another community wide effort to introduce polarizability in the popular CHARMM family of force fields.]

63. Huang J, MacKerell ADJ. Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (AAQAA)(3) Peptide. Biophys. J. 2014; 107:991–997. [PubMed: 25140435]

64. Cerutti DS, Swope WC, Rice JE, Case DA. ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. J Chem Theory Comput. 2014; 10:4515–4534. [PubMed: 25328495]

65. Zgarbová M, Javier Luque F, Sponer J, Otyepka M, Jurecka P. A Novel Approach for Deriving Force Field Torsion Angle Parameters Accounting for Conformation-Dependent Solvation Effects. J Chem Theory Comput. 2012; 8:3232–3242. [PubMed: 26605732]

66. Shirts M, Pande VS. Screen Savers of the World Unite! Science. 2000; 290:1903–1904. [PubMed: 17742054]

67. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P, Shukla D, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. J Chem Theory Comput. 2013; 9:461–469. [PubMed: 23316124]

68. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013; 29:btt055–854.

69. Okazaki K-I, Hummer G. Phosphate release coupled to rotary motion of F1-ATPase. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:16468–16473. [PubMed: 24062450]

70. Ma W, Schulten K. Mechanism of Substrate Translocation by a Ring-Shaped ATPase Motor at Millisecond Resolution. J. Am. Chem. Soc. 2015; 137:3031–3040. [PubMed: 25646698]

71. Okazaki K-I, Hummer G. Elasticity, friction, and pathway of γ-subunit rotation in FoF1-ATP synthase. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:10720–10725. [PubMed: 26261344]

72. Morriss-Andrews A, Shea J-E. Computational studies of protein aggregation: methods and applications. Annu. Rev. Phys. Chem. 2015; 66:643–666. [PubMed: 25648485]

73. Feig M, Harada R, Mori T, Yu I, Takahashi K, Sugita Y. Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. Journal of Molecular Graphics and Modelling. 2015; 58:1–9. [PubMed: 25765281]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

Molecular dynamics is reaching longer timescales faster than Moore's law's rate.

Improved sampling algorithms, force fields and computers drive advances.

At current rate of advance, many single domain proteins will be foldable in 5 years.

Computed ligand binding affinities can be accurately compared with experiment.

Force fields have improved to make prediction, and not only postdiction, possible.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

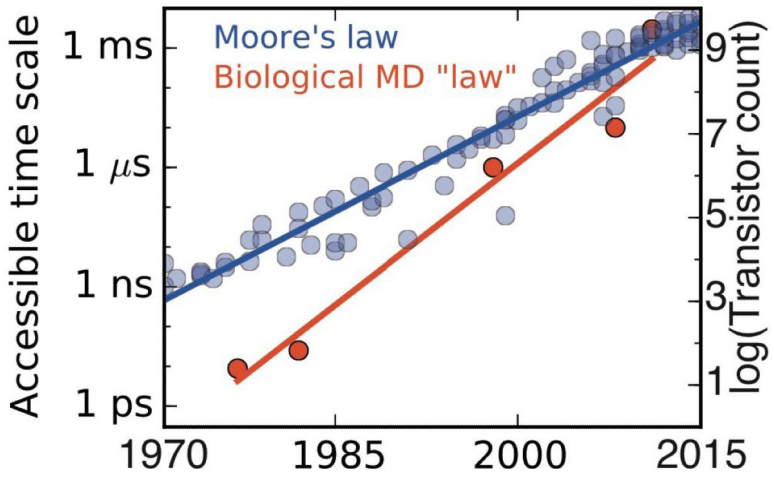

Figure 1.

The accessible time scale for computational biology has grown faster than Moore's law of semiconductors and computing. [1] [2-6].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

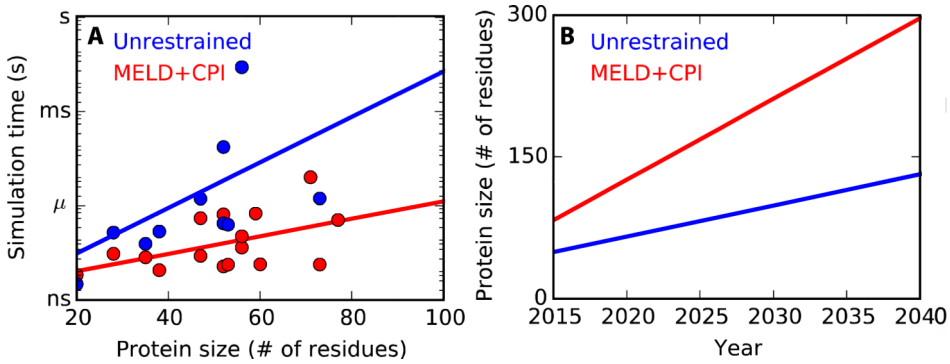

Figure 2.

A. Simulation times for finding native states as a function of protein size, by unrestrained MD, and by MELD, which uses generic directives as restraints. Data from [10,15]. Figure 2B. When will MD find protein native structures of size N? Extrapolations by combining the data in Fig 1 and 2A. The slope for the two lines is calculated as the ratio of the slopes for the biological “law” and the slope for increase in simulation time with protein size. The ordinate in the origin is set to be what simulations can accomplish as of today starting from fully extended chains.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

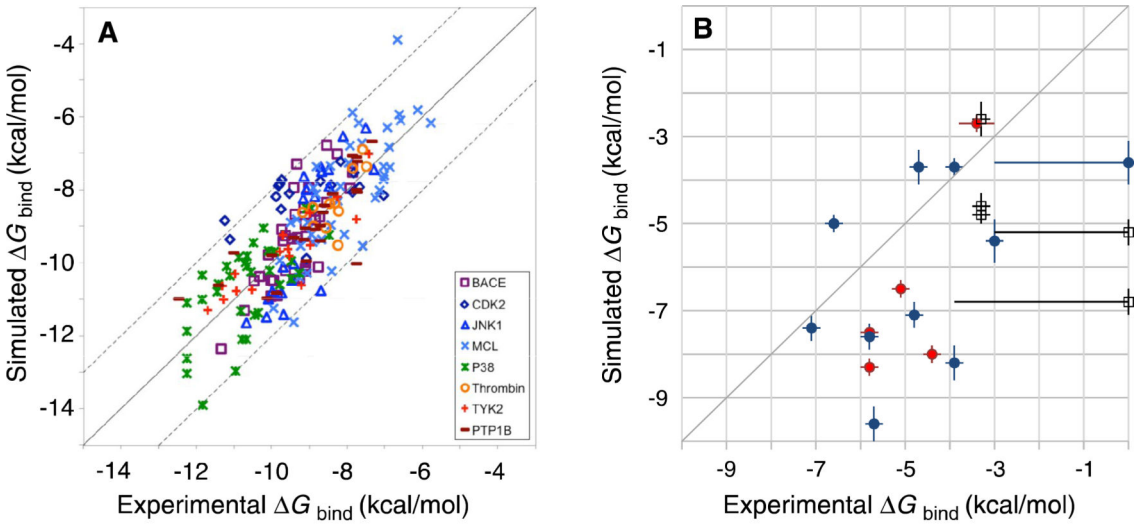

Figure 3. Binding free energy calculations correlate with experiments

Computed RBFEs [34] (A) and computed ABFEs [35] (B) compared to experimental values. While the RMS errors are still relatively large (around 2 kcal/mol), nevertheless, it shows that it is becoming possible to calculate approximate binding affinities. [39]