Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics

Papaleo, Elena

frontiers
in Molecular Biosciences

Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity

Elena Papaleo *

Structural Biology and Nuclear Magnetic Resonance Laboratory, Department of Biology, University of Copenhagen, Copenhagen, Denmark

In the last years, we have been observing remarkable improvements in the field of protein dynamics. Indeed, we can now study protein dynamics in atomistic details over several timescales with a rich portfolio of experimental and computational techniques. On one side, this provides us with the possibility to validate simulation methods and physical models against a broad range of experimental observables. On the other side, it also allows a complementary and comprehensive view on protein structure and dynamics. What is needed now is a better understanding of the link between the dynamic properties that we observe and the functional properties of these important cellular machines. To make progresses in this direction, we need to improve the physical models used to describe proteins and solvent in molecular dynamics, as well as to strengthen the integration of experiments and simulations to overcome their own limitations. Moreover, now that we have the means to study protein dynamics in great details, we need new tools to understand the information embedded in the protein ensembles and in their dynamic signature. With this aim in mind, we should enrich the current tools for analysis of biomolecular simulations with attention to the effects that can be propagated over long distances and are often associated to important biological functions. In this context, approaches inspired by network analysis can make an important contribution to the analysis of molecular dynamics simulations.

Keywords: protein dynamics, molecular dynamics, protein NMR, enhanced sampling, metadynamics, protein structure

Protein Dynamics and Conformational Changes and their Importance in Biology

It is now well-established that proteins, and biomolecules in general, are highly dynamic systems (Vendruscolo, 2007; Boehr et al., 2009; Zhuravlev and Papoian, 2010; Osawa et al., 2012). We moved, in the last decades, from the structure-function paradigm to the structure-dynamics-function triad, where not only the knowledge of the tertiary or quaternary assemblies is important to understand protein function, but also the dynamical behavior on different timescales (Henzler-Wildman and Kern, 2007; Klepeis et al., 2009; Villali and Kern, 2010; Wand, 2012). In this context, the knowledge of even protein states that account for a small population of the ensemble are important (Baldwin and Kay, 2009). Indeed, proteins undergo conformational changes even in their free and unbound/not modified state. Those minor conformations can resemble functional

states, such as the structure that a protein assumes when it binds to another biological partner. The "perturbations" that affect the populations of the protein ensemble are caused by modifications, mutations, and interactions with other biomolecules and have impact on the related biological mechanisms (Popovych et al., 2006; Manley and Loria, 2012; Motlagh et al., 2012; Tsai and Nussinov, 2014).

An Atom-Scale View on Protein Structure and Dynamics by Molecular Dynamics Simulations

Molecular simulations applied to biology are a powerful technique to integrate experimental studies, as attested by the Nobel Prize in Chemistry awarded to Martin Karplus, Micheal Levitt, and Arieh Warshel in 2013 for "the development of multiscale models for complex chemical systems." Among the large array of simulation techniques that can cover the definition of a multiscale approach, I will here focus on atomistic classical molecular dynamics (MD) approaches. Other approaches that are important component of a multiscale approach are extensively covered in many review articles in the field (such as Rapaport, 1998; Koehl and Levitt, 1999; Schlick, 2002; Warshel, 2003; Snow et al., 2004; Xia and Levitt, 2004; Karplus and Kuriyan, 2005; Sherwood et al., 2008; Dror et al., 2012; Saunders and Voth, 2013; Salvatella, 2014; Field, 2015) and are not the focus of this perspective article.

Thirty-eight years are now passed since the first molecular dynamics (MD) simulation of a biomolecule, i.e., the bovine pancreatic trypsin inhibitor, was published (McCammon et al., 1977) and even if that was a short and in vacuum simulation, that work has a remarkable impact on the way in which we look at biomolecular structures. Thirty-eight years later, MD simulations are widespread tools that even experimentalists use to rationalize or guide experiments and a search in Pubmed with "molecular dynamics simulations" as a keyword provides more than 40,000 entries.

In MD methods, the system obeys to Newton's equation of motion and the interactions between atoms are described using, as physical models, molecular mechanics force fields (for a detailed discussion of the basis of MD see, for instance, Rapaport, 1998; Schlick, 2002; Kukol, 2015). Atomistic MD simulations have a unique strength of providing a description of biomolecules on several timescales from the femtosecond (fs) to the millisecond (ms), without renouncing to an atom-scale view for both the biomolecules and the solvent (Dror et al., 2012). They become over the last decades a useful method to integrate experimental research in structural biology and protein science, providing a "computational microscope" for proteins and their complexes (Dror et al., 2012).

The Importance of the Force Field Parameters: Improvements, Limitations and Experimental Validation

The way in which we describe biomolecules in MD simulations depends on the force field, that we use to describe the system.

Recently, new force fields have been developed that are in good agreement with many NMR-derived parameters, describing dynamics over different timescales (Best et al., 2012; Lindorff-Larsen et al., 2012; Huang and MacKerell, 2013; Reif et al., 2013). Despite the fact that the mathematical functional form of the classical force fields is similar, the new-generation force fields differ in parameters that are associated with a subset of very important torsional angles.

Providing the community with good and accurate force fields is an important task. Indeed, the risk to misinterpret the results from our simulations and, as a consequence, to provide biological models that are not meaningful is high. It is known that the even slight changes in force-field parameters have a large impact on the resulting conformational ensemble, as well as on the capability to reproduce experimental parameters (Lindorff-Larsen et al., 2012; Piana et al., 2014).

MD force fields, even if constantly improving, are still far from being perfect (Piana et al., 2014). Indeed, there are several aspects of protein dynamics that we are not able to simulate with accuracy, or even in a reasonable way. For example, currently employed classical MD force fields have been shown to overestimate salt bridges (Debiec et al., 2014). This observation can change how we interpret effects related to electrostatic interactions in proteins. Salt bridges, especially when they are in solvent-exposed positions, may be less populated in solution than expected from the analysis of MD simulations or X-ray structures, as convincingly argued by the NMR study of GB1 in solution (Tomlinson et al., 2009) or by the observation that mutations of residues involved in salt bridges in thermophilic or hyperthermophilic enzymes have often minor effects on thermal stability (Jónsdóttir et al., 2014). Moreover, classical force fields cannot accurately describe those interactions that need polarizable effects to be explicitly taken into account. This is crucial, for example, when we need to deal with metal-binding proteins (Banci, 2003). MD force fields also encounter the risk of over-compaction of the protein during the simulation, and this is especially important when we aim at studying systems as intrinsically disordered proteins (IDPs) or unfolded states (Knott and Best, 2012; Lambrughi et al., 2012; Invernizzi et al., 2013; Camilloni and Vendruscolo, 2014; Piana et al., 2014; Palazzesi et al., 2015). We encounter the risk, for example, to attribute a great importance to intramolecular interactions observed in MD-derived IDPs ensembles, which are a consequence of the highly compact states sampled during the simulation. Very recently, solutions have been provided for MD simulations of IDPs or unfolded states (Best et al., 2014; Piana et al., 2015). The authors of these works showed that it is possible to recover the compactness of IDPs or unfolded proteins observed, for example, by small-angle X-ray scattering thank to improved solvent models. This is an encouraging step toward a more accurate and reliable atomistic description of IDPs by simulations.

A continuous exchange between experimental biophysical techniques and MD is also needed.

Indeed, evaluation and validation of MD force fields cannot rely only on the study of few proteins, we need to move forward and test force fields on proteins with other folds. To do this, a large amount of experimental data are necessary for the comparison, along with methods to back-calculate

from the simulated ensemble properties that can be measured experimentally, for instance by NMR spectroscopy or SAXS (Lindorff-Larsen et al., 2005; Lange et al., 2008; Kohlhoff et al., 2009; Sahakyan et al., 2011; Li and Brüschweiler, 2012; Camilloni and Vendruscolo, 2014). In this context, the usage of NMR parameters to cross-validate a MD ensemble should become a routine practice in computational biophysics.

The Coverage of the Conformational Space in MD Simulations

On the other side, one of the main well-known issues of classical atomistic MD is the sampling of the conformational space. Indeed, during classical MD simulations, even if now we have the hardware (Friedrichs et al., 2009; Kohlhoff et al., 2014; Shaw et al., 2014) and software (Harvey et al., 2009; Pronk et al., 2013) to improve the performance and simulate even millisecond timescales, the protein encounters the risk to be trapped for a long time in a local basin of the conformational space, making the simulations not converging on even long timescales. Conventional MD simulations thus allow us to provide a limited description of protein dynamics since we neglect a major part of the conformational landscape. Also, we should keep in mind the risk to sample regions of the conformational space that are not relevant for the dynamics that we observe experimentally when a force field in bad agreement with the experimental data is used (Lindorff-Larsen et al., 2012).

The sampling problem has been recently overcome, for example, by enhanced-sampling techniques integrated to the atomistic force field description that MD provides. A rich portfolio of methods for enhancing sampling has been developed, each of them relying on different philosophies (Piana and Laio, 2007; Hritz and Oostenbrink, 2008; Sutto et al., 2012; Abrams and Bussi, 2013; Bernardi et al., 2014; Do et al., 2014; Papaleo et al., 2014; Spiwok et al., 2014; Barducci et al., 2015). Among them, metadynamics has the potential to enhance the sampling of rare transitions, which are often important in biology (Laio and Gervasio, 2008). Metadynamics approaches proved to be very accurate in recovering the conformational changes experimentally observed on the micro-millisecond time scale, which are the most important for many biological processes (Berteotti et al., 2009; Palazzesi et al., 2013; Sutto and Gervasio, 2013; Papaleo et al., 2014). In metadynamics, the sampling is enhanced along a properly selected set of reaction coordinates (i.e., collective variables) that have to account for the slowest degrees of freedom of the process of interest. It can be argued that it is limited to the usage of few collective variables at the time to ensure convergent simulations. The community, however, did extensive efforts in the last years to mitigate this problem, for instance, by reweighting procedures to calculate collective variables not used directly to bias the simulations (Bonomi et al., 2009) or by approaches as bias exchange metadynamics (Piana and Laio, 2007), in which the replicas exchange between different collective variables rather than in the temperature space.

The development of methods to predict NMR-derived parameters that are probes of dynamics over different timescales

mentioned in the previous section, not only provides a solid strategy to evaluate the quality of MD ensembles but it also allow to simulate the protein integrating the atomistic description of the MD force fields with the experimental data. In this context, the usage of NMR parameters to restrain MD simulations is becoming a popular application and provided encouraging results toward a more complete picture of the complex dynamics of proteins (Lindorff-Larsen et al., 2005; Tang et al., 2007; Vendruscolo, 2007; Fenwick et al., 2011; Camilloni et al., 2012, 2013; Camilloni and Vendruscolo, 2014). In some of these approaches, the restraints are generally averaged over multiple replicas to better reflect the intrinsic average nature of the experimental data. These computational methods, among the enhanced sampling techniques that can be integrated to the atomistic MD force fields, have also the capability to recover the dynamics observed experimentally in the millisecond regime with high accuracy and without limiting the analysis to a small subset of collective variables, as metadynamics does.

When the Structural Effects Come from Distal Site: Long-Range Communication in Protein Dynamics and Network Analysis Applied to MD Simulations

The network paradigm has been extensively used to describe structure, topology and dynamics of proteins (Vishveshwara et al., 2009; Atilgan et al., 2012; Collier and Ortiz, 2013; Csermely et al., 2013; Feher et al., 2014). The intramolecular non-covalent interactions in a protein are known to be crucial in determining the protein structure and they can be collectively represented in the form of a network, namely a Protein Structure Network (PSN), where the residues are the nodes of the network connected by edges that depend on their interaction strength. PSNs are "small worlds" (Vendruscolo et al., 2002; Atilgan et al., 2004) suitable for the fast transmission of conformational changes at distal sites. Indeed, in the small world of PSNs, the residues can communicate through the shortest paths available and multiple paths with common nodes are used (Del Sol et al., 2009). Several important issues, however, are, still unsolved when it comes to network analysis of protein ensembles. We can identify paths of communications between distal residues and we see that they are modified by perturbations. Very often, however, the way in which the native network of a protein is affected by a perturbation is not easy to predict. We do not exactly know if the paths that we calculate in a structural ensemble are really relevant for functional dynamics, or even for the conformational changes observed in the experiments. The community is collecting some examples in favor of a relationship between the networks observed in the protein structure and functional properties of proteins (Whitley and Lee, 2009; Mariani et al., 2013; Van den Bedem et al., 2013; Invernizzi et al., 2014; Papaleo et al., 2012b, 2014). We are, however, still far from being able to derive straightforward information and predictions out of the protein network. More efforts are certainly needed in this direction, also considering the potential that these approaches can have in

applications such as drug discovery and protein engineering (Kar et al., 2010; Nussinov and Tsai, 2013).

Several methods are now available to identify networks and paths of communication in protein ensembles, such as the ones collected by MD simulations. The originally named PSN approach relies on the description of atomic contacts between residues that also feature correlated motions (Ghosh and Vishveshwara, 2007; Seeber et al., 2011; Papaleo et al., 2012a). It has become very popular also thanks to the availability of tools to derive PSN from both experimental and simulated ensembles (Seeber et al., 2011; Pasi et al., 2012; Bhattacharyya et al., 2013; Tiberti et al., 2014). The metrics to estimate correlated pairs of residues can range from Person correlation to Mutual Information and often suffer of the disadvantage in terms of convergence and statistical significance of the data. Caution has thus to be taken in deriving correlation maps from a MD ensemble and we should avoid in MD applications to biological target to use just a unique average correlation map from an entire MD trajectory. Other methods have been recently proposed and applied to MD simulations, relying on different philosophies, such as the usage of structural alphabets (Pandini et al., 2013), the integration of topology, correlated motions and communication propensity (Allain et al., 2014), the notion of energetic coupling between protein residues. (Ribeiro and Ortiz, 2014) or even the application of concepts inspired by engineering in which the protein is treated as a mechanical construct rather than a chemical molecule (Stacklies et al., 2011) with the possibility to trace communication even in stiff structures without evident atomic displacement.

The contribution of the solvent in the modulation of long-range communication in proteins should not be neglected and the community still lacks successful approaches to analyze the solvent-protein intermolecular networks, making this an appealing field for future research.

All the methods mentioned above often rely on the integration between network analysis and MD simulations. It is thus important to achieve a proper conformational sampling during the simulations, verify the reproducibility of the results over different replicates of the same system and o evaluate how long the simulations have to be extended to get meaningful results from network analysis. Indeed, long-range effects are often associated to conformational changes occurring on microsecond-millisecond timescales. Thus, the limitation inherent in the sampling achieved by classical MD has to be kept in mind. We recently showed that MD simulations of 100 ns or 1 μs of the same system provided similar average PSN results (Invernizzi et al., 2014) but this account only for one case of study, making any generalization very hard.

The Strength is in Unity: Integration of Molecular Dynamics to Experimental Observable

In conclusion, we have reached now a point in which both from the computational and the experimental side, there are a large amount of information on protein dynamics and techniques to explore it on different timescale, as well as techniques to disclose effects that can be mediated by distal communicating sites and to actually describe these paths of communication at the atom-scale. The strength is now in the unity, in the capability to tightly integrate the computational and experimental approaches combining their own capabilities and overcoming their own limitations for a more detailed and accurate understanding of protein dynamics at the atomic level.

References

Abrams, C., and Bussi, G. (2013). Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* 16, 163–199. doi: 10.3390/e16010163

Allain, A., Chauvot de Beauchêne, I., Langenfeld, F., Guarracino, Y., Laine, E., and Tchertanov, L. (2014). Allosteric pathway identification through network analysis: from molecular dynamics simulations to interactive 2D and 3D graphs. *Faraday Discuss.* 169, 303–321. doi: 10.1039/C4FD00024B

Atilgan, A. R., Akan, P., and Baysal, C. (2004). Small-world communication of residues and significance for protein dynamics. *Biophys. J.* 86, 85–91. doi: 10.1016/S0006-3495(04)74086-2

Atilgan, C., Okan, O. B., and Atilgan, A. R. (2012). Network-based models as tools hinting at nonevident protein functionality. *Annu. Rev. Biophys.* 41, 205–225. doi: 10.1146/annurev-biophys-050511-102305

Baldwin, A. J., and Kay, L. E. (2009). NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* 5, 808–814. doi: 10.1038/nchembio.238

Banci, L. (2003). Molecular dynamics simulations of metalloproteins. *Curr. Opin. Chem. Biol.* 7, 143–149. doi: 10.1016/S1367-5931(02)00014-5

Barducci, A., Pfaendtner, J., and Bonomi, M. (2015). Tackling sampling challenges in biomolecular simulations. *Methods Mol. Biol.* 1215, 151–171. doi: 10.1007/978-1-4939-1465-4_8

Bernardi, R. C., Melo, M. C. R., and Schulten, K. (2014). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* 1850, 872–877. doi: 10.1016/j.bbagen.2014.10.019

Berteotti, A., Cavalli, A., Branduardi, D., Gervasio, F. L., Recanatini, M., and Parrinello, M. (2009). Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. *J. Am. Chem. Soc.* 131, 244–250. doi: 10.1021/ja806846q

Best, R. B., Zheng, W., and Mittal, J. (2014). Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* 10, 5113–5124. doi: 10.1021/ct500569b

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x

Bhattacharyya, M., Bhat, C. R., and Vishveshwara, S. (2013). An automated approach to network features of protein structure ensembles. *Protein Sci.* 22, 1399–1416. doi: 10.1002/pro.2333

Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5, 789–796. doi: 10.1038/nchembio.232

Bonomi, M., Barducci, A., and Parrinello, M. (2009). Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* 30, 1615–1621. doi: 10.1002/jcc.21305

Camilloni, C., Cavalli, A., and Vendruscolo, M. (2013). Replica-averaged metadynamics. *J. Chem. Theory Comput.* 9, 5610–5617. doi: 10.1021/ct4006272

Camilloni, C., Robustelli, P., De Simone, A., Cavalli, A., and Vendruscolo, M. (2012). Characterization of the conformational equilibrium between the two

major substates of RNase A using NMR chemical shifts. *J. Am. Chem. Soc.* 134, 3968–3971. doi: 10.1021/ja210951z

Camilloni, C., and Vendruscolo, M. (2014). Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc.* 136, 8982–8991. doi: 10.1021/ja5027584

Collier, G., and Ortiz, V. (2013). Emerging computational approaches for the study of protein allostery. *Arch. Biochem. Biophys.* 538, 6–15. doi: 10.1016/j.abb.2013.07.025

Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138, 333–408. doi: 10.1016/j.pharmthera.2013.01.016

Debiec, K. T., Gronenborn, A. M., Chong, L. T., and Fields, F. (2014). Evaluating the strength of salt bridges – a comparison of current biomolecular force fields. *J. Phys. Chem. B,* 118, 6561–6569. doi: 10.1021/jp500958r

Del Sol, A., Tsai, C.-J., Ma, B., and Nussinov, R. (2009). The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* 17, 1042–1150. doi: 10.1016/j.str.2009.06.008

Do, T. N., Choy, W.-Y., and Karttunen, M. (2014). Accelerating the conformational sampling of intrinsically disordered proteins. *J. Chem. Theory Comput.* 10, 5081–5094. doi: 10.1021/ct5004803

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* 41, 429–452. doi: 10.1146/annurev-biophys-042910-155245

Feher, V. A., Durrant, J. D., Van Wart, A. T., and Amaro, R. E. (2014). Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* 25, 98–103. doi: 10.1016/j.sbi.2014.02.004

Fenwick, R. B., Esteban-Martín, S., Richter, B., Lee, D., Walter, K. F. A., Milovanovic, D., et al. (2011). Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.* 133, 10336–10339. doi: 10.1021/ja200461n

Field, M. (2015). Technical advances in molecular simulation since the 1980s. *Arch. Biochem. Biophys.* doi: 10.1016/j.abb.2015.03.005. [Epub ahead of print].

Friedrichs, M. S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A. L., et al. (2009). Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30, 864–872. doi: 10.1002/jcc.21209

Ghosh, A., and Vishveshwara, S. (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15711–15716. doi: 10.1073/pnas.0704459104

Harvey, M. J., Giupponi, G., and Fabritiis, G. De. (2009). ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* 5, 1632–1639. doi: 10.1021/ct9000685

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522

Hritz, J., and Oostenbrink, C. (2008). Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* 128, 144121. doi: 10.1063/1.2888998

Huang, J., and MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* 34, 2135–2145. doi: 10.1002/jcc.23354

Invernizzi, G., Lambrughi, M., Regonesi, M. E., Tortora, P., and Papaleo, E. (2013). The conformational ensemble of the disordered and aggregation-protective 182-291 region of ataxin-3. *Biochim. Biophys. Acta* 1830, 5236–5247. doi: 10.1016/j.bbagen.2013.07.007

Invernizzi, G., Tiberti, M., Lambrughi, M., Lindorff-Larsen, K., and Papaleo, E. (2014). Communication routes in ARID domains between distal residues in helix 5 and the DNA-binding loops. *PLoS Comput. Biol.* 10:e1003744. doi: 10.1371/journal.pcbi.1003744

Jónsdóttir, L. B., Ellertsson, B. Ö., Invernizzi, G., Magnúsdóttir, M., Thorbjarnardóttir, S. H., Papaleo, E., et al. (2014). The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase. *Biochim. Biophys. Acta* 1844, 2174–2181. doi: 10.1016/j.bbapap.2014.08.011

Kar, G., Keskin, O., Gursoy, A., and Nussinov, R. (2010). Allostery and population shift in drug discovery. *Curr. Opin. Pharmacol.* 10, 715–722. doi: 10.1016/j.coph.2010.09.002

Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6679–6685. doi: 10.1073/pnas.0408930102

Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19, 120–127. doi: 10.1016/j.sbi.2009.03.004

Knott, M., and Best, R. (2012). A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. *PLoS Comput. Biol.* 8:e1002605. doi: 10.1371/journal.pcbi.1002605

Koehl, P., and Levitt, M. (1999). Theory and simulation. Can theory challenge experiments. *Curr. Opin. Struct. Biol.* 9, 155–156. doi: 10.1016/S0959-440X(99)80021-9

Kohlhoff, K. J., Robustelli, P., Cavalli, A., Salvatella, X., and Vendruscolo, M. (2009). Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* 131, 13894–13895. doi: 10.1021/ja903772t

Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., et al. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* 6, 15–21. doi: 10.1038/nchem.1821

Kukol, A. (2015). *Molecular Modeling of Proteins*. New York, NY: Springer.

Laio, A., and Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* 71:126601. doi: 10.1088/0034-4885/71/12/126601

Lambrughi, M., Papaleo, E., Testa, L., Brocca, S., De Gioia, L., and Grandori, R. (2012). Intramolecular interactions stabilizing compact conformations of the intrinsically disordered kinase-inhibitor domain of Sic1: a molecular dynamics investigation. *Front. Physiol.* 3:435. doi: 10.3389/fphys.2012.00435

Lange, O. F., Lakomek, N.-A., Farès, C., Schröder, G. F., Walter, K. F. A., Becker, S., et al. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320, 1471–1475. doi: 10.1126/science.1157092

Li, D.-W., and Brüschweiler, R. (2012). PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *J. Biomol. NMR* 54, 257–265. doi: 10.1007/s10858-012-9668-8

Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* 433, 128–132. doi: 10.1038/nature03199

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PLoS ONE* 7:e32131. doi: 10.1371/journal.pone.0032131

Manley, G., and Loria, J. P. (2012). NMR insights into protein allostery. *Arch. Biochem. Biophys.* 519, 223–231. doi: 10.1016/j.abb.2011.10.023

Mariani, S., Dell'Orco, D., Felline, A., Raimondi, F., and Fanelli, F. (2013). Network and atomistic simulations unveil the structural determinants of mutations linked to retinal diseases. *PLoS Comput. Biol.* 9:e1003207. doi: 10.1371/journal.pcbi.1003207

McCammon, J., Gelin, B., and Karplus, M. (1977). Dynamics of folded proteins. *Nature* 267, 585–590. doi: 10.1038/267585a0

Motlagh, H. N., Li, J., Thompson, E. B., and Hilser, V. J. (2012). Interplay between allostery and intrinsic disorder in an ensemble. *Biochem. Soc. Trans.* 40, 975–980. doi: 10.1042/BST20120163

Nussinov, R., and Tsai, C.-J. (2013). Allostery in disease and in drug discovery. *Cell* 153, 293–305. doi: 10.1016/j.cell.2013.03.034

Osawa, M., Takeuchi, K., Ueda, T., Nishida, N., and Shimada, I. (2012). Functional dynamics of proteins revealed by solution NMR. *Curr. Opin. Struct. Biol.* 22, 660–669. doi: 10.1016/j.sbi.2012.08.007

Palazzesi, F., Barducci, A., Tollinger, M., and Parrinello, M. (2013). The allosteric communication pathways in KIX domain of CBP. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14237–14242. doi: 10.1073/pnas.1313548110

Palazzesi, F., Prakash, M. K., Bonomi, M., and Barducci, A. (2015). Accuracy of current all-atom force-fields in modeling protein disordered states. *J. Chem. Theory Comput.* 11, 2–7. doi: 10.1021/ct500718s

Pandini, A., Fornili, A., Fraternali, F., and Kleinjung, J. (2013). GSATools: analysis of allosteric communication and functional local motions using a structural alphabet. *Bioinformatics* 29, 2053–2055. doi: 10.1093/bioinformatics/btt326

Papaleo, E., Lindorff-Larsen, K., and De Gioia, L. (2012a). Paths of long-range communication in the E2 enzymes of family 3: a molecular

dynamics investigation. *Phys. Chem. Chem. Phys.* 14, 12515–12525. doi: 10.1039/c2cp41224a

Papaleo, E., Renzetti, G., and Tiberti, M. (2012b). Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: a molecular dynamics investigation. *PLoS ONE* 7:e35686. doi: 10.1371/journal.pone.0035686

Papaleo, E., Sutto, L., Gervasio, F. L., and Lindorff-Larsen, K. (2014). Conformational changes and free energies in a proline isomerase. *J. Chem. Theory Comput.* 10, 4169–4174. doi: 10.1021/ct500536r

Pasi, M., Tiberti, M., Arrigoni, A., and Papaleo, E. (2012). xPyder: a PyMOL plugin to analyze coupled residues and their networks in protein structures. *J. Chem. Inf. Model.* 52, 1865–1874. doi: 10.1021/ci300213c

Piana, S., Donchev, A., Robustelli, P., and Shaw, D. (2015). Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* 119, 5113–5123. doi: 10.1021/jp508971m

Piana, S., Klepeis, J. L., and Shaw, D. E. (2014). Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 24C, 98–105. doi: 10.1016/j.sbi.2013.12.006

Piana, S., and Laio, A. (2007). A bias-exchange approach to protein folding. *J. Phys. Chem. B* 111, 4553–4559. doi: 10.1021/jp067873l

Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006). Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13, 831–838. doi: 10.1038/nsmb1132

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/bioinformatics/btt055

Rapaport, D. C. (1998). *The Art of Molecular Dynamics Simulation*. Cambridge, UK: Cambridge University Press.

Reif, M. M., Winger, M., and Oostenbrink, C. (2013). Testing of the GROMOS force-field parameter set 54A8: structural properties of electrolyte solutions, lipid bilayers, and proteins. *J. Chem. Theory Comput.* 9, 1247–1264. doi: 10.1021/ct300874c

Ribeiro, A. A. S. T., and Ortiz, V. (2014). Determination of signaling pathways in proteins through network theory: importance of the topology. *J. Chem. Theory Comput.* 10, 1762–1769. doi: 10.1021/ct400977r

Sahakyan, A. B., Vranken, W. F., Cavalli, A., and Vendruscolo, M. (2011). Structure-based prediction of methyl chemical shifts in proteins. *J. Biomol. NMR* 50, 331–346. doi: 10.1007/s10858-011-9524-2

Salvatella, X. (2014). Understanding protein dynamics using conformational ensembles. *Adv. Exp. Med. Biol.* 805, 67–85. doi: 10.1007/978-3-319-02970-2_3

Saunders, M., and Voth, G. (2013). Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* 42, 73–93. doi: 10.1146/annurev-biophys-083012-130348

Schlick, T. (2002). *Molecular Modeling and Simulation: An Interdisciplinary Guide*. New York, NY: Springer.

Seeber, M., Felline, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., et al. (2011). Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* 32, 1183–1194. doi: 10.1002/jcc.21688

Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., et al. (2014). "Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (New Orleans, LA: IEEE Press), 41–53.

Sherwood, P., Brooks, B., and Sansom, M. (2008). Multiscale methods for macromolecular simulations. *Curr. Opin. Struct. Biol.* 18, 630–640. doi: 10.1016/j.sbi.2008.07.003

Snow, C., Sorin, E., Rhee, Y., and Pande, V. (2004). How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34, 43–69. doi: 10.1146/annurev.biophys.34.040204.144447

Spiwok, V., Suæur, Z., and Hošek, P. (2014). Enhanced sampling techniques in biomolecular simulations. *Biotechnol. Adv.* doi: 10.1016/j.biotechadv.2014.11.011. [Epub ahead of print].

Stacklies, W., Seifert, C., and Graeter, F. (2011). Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinformatics* 12:101. doi: 10.1186/1471-2105-12-101

Sutto, L., and Gervasio, F. L. (2013). Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10616–10621. doi: 10.1073/pnas.1221953110

Sutto, L., Marsili, S., and Gervasio, F. L. (2012). New advances in metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2, 771–779. doi: 10.1002/wcms.1103

Tang, C., Schwieters, C. D., and Clore, G. M. (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449, 1078–1082. doi: 10.1038/nature06232

Tiberti, M., Invernizzi, G., Lambrughi, M., Inbar, Y., Schreiber, G., and Papaleo, E. (2014). PyInteraph: a framework for the analysis of interaction networks in structural ensembles of proteins. *J. Chem. Inf. Model.* 54, 1537–1551. doi: 10.1021/ci400639r

Tomlinson, J. H., Ullah, S., Hansen, P. E., and Williamson, M. P. (2009). Characterization of salt bridges to lysines in the protein G B1 domain. *J. Am. Chem. Soc.* 131, 4674–4684. doi: 10.1021/ja808223p

Tsai, C.-J., and Nussinov, R. (2014). A unified view of "how allostery works". *PLoS Comput. Biol.* 10:e1003394. doi: 10.1371/journal.pcbi.1003394

Van den Bedem, H., Bhabha, G., Yang, K., Wright, P. E., and Fraser, J. S. (2013). Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* 10, 896–902. doi: 10.1038/nmeth.2592

Vendruscolo, M. (2007). Determination of conformationally heterogeneous states of proteins. *Curr. Opin. Struct. Biol.* 17, 15–20. doi: 10.1016/j.sbi.2007.01.002

Vendruscolo, M., Dokholyan, N., Paci, E., and Karplus, M. (2002). Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65, 1–4. doi: 10.1103/PhysRevE.65.061910

Villali, J., and Kern, D. (2010). Choreographing an enzyme's dance. *Curr. Opin. Chem. Biol.* 14, 636–643. doi: 10.1016/j.cbpa.2010.08.007

Vishveshwara, S., Ghosh, A., and Hansia, P. (2009). Intra and inter-molecular communications through protein structure network. *Curr. Protein Pept. Sci.* 10, 146–160. doi: 10.2174/138920309787847590

Wand, A. J. (2012). The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Biol.* 23, 75–81. doi: 10.1016/j.sbi.2012.11.005

Warshel, A. (2003). Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.* 32, 425–443. doi: 10.1146/annurev.biophys.32.110601.141807

Whitley, M. J., and Lee, A. L. (2009). Frameworks for understanding long-range intra-protein communication. *Curr. Protein Pept. Sci.* 10, 116–127. doi: 10.2174/138920309787847563

Xia, Y., and Levitt, M. (2004). Simulating protein evolution in sequence and structure space. *Curr. Opin. Struct. Biol.* 14, 202–207. doi: 10.1016/j.sbi.2004.03.001

Zhuravlev, P. I., and Papoian, G. A. (2010). Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q. Rev. Biophys.* 43, 295–332. doi: 10.1017/S0033583510000119