

Computational Tool

MDN: A Web Portal for Network Analysis of Molecular Dynamics Simulations

Andre A. S. T. Ribeiro¹ and Vanessa Ortiz^{1,*}

¹Department of Chemical Engineering, Columbia University, New York, New York

ABSTRACT We introduce a web portal that employs network theory for the analysis of trajectories from molecular dynamics simulations. Users can create protein energy networks following methodology previously introduced by our group, and can identify residues that are important for signal propagation, as well as measure the efficiency of signal propagation by calculating the network coupling. This tool, called MDN, was used to characterize signal propagation in *Escherichia coli* heat-shock protein 70-kDa. Two variants of this protein experimentally shown to be allosterically active exhibit higher network coupling relative to that of two inactive variants. In addition, calculations of partial coupling suggest that this quantity could be used as part of the criteria to determine pocket druggability in drug discovery studies.

INTRODUCTION

Computer simulations have become a widely used tool to explore the behavior of molecular systems. Great advances have been achieved in the last decade, and improvements in both software and hardware allow for calculations of ever-increasing length and complexity to be performed (1,2). However, performing a simulation is arguably the least interesting and complex part of a computational study of a molecular system. The raw output of such simulations is typically a trajectory file, which contains 3D coordinates for all the atoms in the simulation box, for different conformations explored by the system during the course of the simulation. For simulations of biomolecular systems, the number of atoms ($\sim 10^5 - 10^7$) coupled to the significant sampling requirements ($\sim 10^7 - 10^{10}$ simulation steps) lead to an enormous amount of high-dimensional data output that needs to be analyzed. This analysis step is thus the core of the investigation itself. Provided the simulations were carried out appropriately, it is the analysis of these data that will determine the quality of the resulting scientific conclusions.

A number of software packages are available for performing molecular dynamics (MD) simulations of biomolecules. The most well known are arguably AMBER (3), CHARMM (4), GROMACS (5), GROMOS (6), and NAMD (7). These have slightly different functionality and provide their own set of analysis tools, ranging from simple calculations of average atomic positions to more complex tasks such as clustering and principal-component analysis (8).

A particular application that computational studies have frequently pursued is the investigation of allostery and protein signaling at the microscopic level (9). A number of

specialized methods have been proposed to address this issue (10–15). Another approach that has been widely used in recent years is to analyze standard MD simulations with network theory (16–22). This approach involves modeling the protein as a graph, with residues mapped to network nodes, thus yielding a coarse-grained representation of the system. The most important choice that needs to be made is how to assign the network connections (*edges*) and the corresponding weights. This is usually done based on interresidue distances and correlation coefficients extracted from MD simulations. We have recently shown that this approach introduces significant statistical errors in the analysis, and networks constructed in this way were not able to identify residues important for signal propagation in the protein imidazole glycerol phosphate synthase (21) or discriminate allosterically active mutants of the lactose repressor protein (22). On the other hand, we showed that assigning network edges based on interaction energies extracted from MD simulations can give an accurate description of signal propagation in proteins (21). In addition, we introduced a network coupling measure that, when applied to protein energy networks, was able to discriminate allosterically active mutants of the lac repressor protein (22). These analyses were done with a customized version of the GROMACS software (5), along with several programs written entirely by our group. Here, we introduce a web portal that enables the performance of this type of network analysis in a user-friendly manner. The portal, called MDN, is available at <http://mdn.chem.columbia.edu>, and it can currently be used to analyze output from simulations performed with the GROMACS and NAMD packages. In addition, we present results obtained by using the server to analyze molecular dynamics simulations of four variants of the *Escherichia coli* heat-shock protein 70 kDa (Hsp70) (DnaK). We show that the energetic

Submitted February 24, 2015, and accepted for publication June 3, 2015.

*Correspondence: vortiz@columbia.edu

Editor: Ivet Bahar.

© 2015 by the Biophysical Society
0006-3495/15/09/1110/7

<http://dx.doi.org/10.1016/j.bpj.2015.06.013>



coupling analysis is able to discriminate allosterically active from inactive mutants, in accordance with previous experimental results. Finally, we calculated partial coupling values and showed that this quantity can help predict the response of a protein to drug binding to a specific pocket in the molecule.

MATERIALS AND METHODS

Software routines for network analysis and server details

The protocol for the construction of protein energy networks involves the calculation of a matrix of interresidue interaction energies. This is accomplished by reading the trajectory file with the corresponding simulation program and calculating the energies without integrating the equations of motion. The protocol is currently available for the programs GROMACS (5) and NAMD (7). Here, we will focus on analyzing a simulation performed with GROMACS. Instructions for NAMD users can be found on the MDN web site. To setup the GROMACS simulation for calculating interaction energies, energy groups need to be defined and explicitly included in the GROMACS simulation parameter file (.mdp). However, the official GROMACS release limits the number of energy groups used in a single simulation. We made small modifications to the GROMACS source code to circumvent this limitation, and to only write to disk the final (average) energies, thus avoiding unnecessary file input/output during post-processing steps. We made similar modifications to the NAMD source code.

Technical details on the implementation of the analysis tool (programming languages used, data management, etc.) are described in the MDN documentation, and are available to the public on the MDN web site.

Hsp70 simulations

We performed MD simulations of several variants of DnaK. The crystal structure of DnaK (23) (Protein Data Bank (PDB) entry 2KHO) was used as the starting point for the wild-type (WT), D388R, D393R, and K70E simulations. The mutants were prepared with the software UCSF Chimera (24). Structures were then solvated with TIP3P water molecules (25) in a truncated octahedral box with minimum distance between the box edges and any protein atoms set to 10 Å. The AMBER-03 force field was used to describe the protein systems (26). Long-range electrostatic interactions were treated with the particle-mesh Ewald method, with a cut-off of 12 Å, whereas van der Waals interactions were switched off between 10 and 12 Å. The systems were minimized with the steepest-descent algorithm, followed by constant-temperature MD simulations at 300 K for 500 ps and equilibration runs at $T = 300$ K and $p = 1$ bar for a further nanosecond. Finally, production runs of 200 ns were performed. Version 4.6.3 of the GROMACS simulation package (5) was used to perform the simulations. The thermostat of Bussi and co-workers (27) was used to maintain the temperature of the systems, and pressure control was achieved with the Parrinello-Rahman algorithm (28). An integration time step of 2 fs was used in the MD simulations, with bond lengths of protein and water molecules constrained with the LINCS (29) and SETTLE (30) algorithms, respectively.

RESULTS AND DISCUSSION

Available types of analysis

The basic input for MDN consists of a trajectory file, as well as the files used for generating the simulation input. These are processed to construct the network topology and extract

the interaction energies to assign network edges, following methodology previously introduced (21). Briefly, weights ω_{ij} are assigned to different residue pairs as follows:

$$\omega_{ij} = \begin{cases} \omega_b, & \text{if } i \text{ and } j \text{ are covalently bound} \\ \chi_{ij}, & \text{otherwise.} \end{cases} \quad (1)$$

The function χ_{ij} is evaluated by calculating average interaction energies, ϵ_{ij} . We then define $\chi_{ij} \equiv 0.5 \times \{1 - (\epsilon_{ij} - \epsilon_{av}) / (5 \times \epsilon_{\text{rmsd}})\}$, where ϵ_{av} and ϵ_{rmsd} are the average and root mean-square deviation of the distribution of interaction energies. The parameter ω_b is set to 0.99, the maximum weight allowed. The resulting network is analyzed to determine nodes important for signaling and the efficiency of signal propagation.

The node betweenness is a measure of network centrality suitable for determining important nodes for signaling (21,31). This quantity measures the number of pathways that go through each node in the network. The Floyd-Warshall algorithm (32) is used to calculate the shortest pathways between all pairs of nodes specified by the user, and the resulting node betweenness and normalized node betweenness, η , are calculated according to the methods of Ribeiro and Ortiz (21).

The network coupling, recently introduced by our group (22), is a measure of the efficiency of signal propagation through the network. It uses a network efficiency measure previously introduced by Latora and Marchiori (33):

$$\epsilon \equiv \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \quad (2)$$

where N is the number of nodes and d_{ij} is the minimum-path distance between nodes i and j . The network coupling is then determined by using Eq. 2 to calculate the efficiencies of the constructed network and of the corresponding ideal network (22):

$$\text{Coupling} = \frac{\epsilon}{\epsilon_{\text{ideal}}} \quad (3)$$

The ideal network uses the pathways calculated for the analyzed network but gives maximum weight to all edges along these pathways (22). We note that Eq. 3 can also be used to calculate partial energetic coupling values between different regions of the network by summing only over the pathways connecting the corresponding groups of nodes.

Job registration and file upload

The start page of MDN exhibits a registration form in which the name and email of the user should be provided. A verification email is sent to the user, who can then choose the simulation package and start to set up the analysis job. Before network construction can be attempted, a number of files need to be uploaded to the server.

The files needed for analyzing a simulation performed with GROMACS are the trajectory file (.xtc, .trr, .gro, or .pdb), and all files used to set up the simulation with the grompp tool. These include the input configuration file (.gro), the simulation parameter file (.mdp), the index file (.ndx), and the system topology file (.top). The topology file usually reads auxiliary .itp files if there is more than one biomolecule in the system. MDN reads the topology file and asks for the upload of the corresponding auxiliary files. After all necessary files are uploaded, the server checks them for consistency and provides a short summary of the system, as can be seen in Fig. 1 for the DnaK test case. We note that MDN enforces a maximum size for file uploads, currently set to 100 MB. Users may remove solvent molecules from their trajectories before upload, resulting in significantly reduced file sizes. Instructions on how to accomplish this task can be found in the MDN documentation. We also note that this method was found to be very robust with regard to structural fluctuations (21), and the user should be able to upload a trajectory file containing a relatively small number of frames, provided no major conformational changes are observed.

File	Name	
Configuration	input.gro	
Topology	2kho.top	
MDP	md.mdp	
Trajectory	trajectory.xtc	
Index	domain.ndx	
forcefield.itp	forcefield.itp	
2kho.top	2kho.top	
posre.itp	posre.itp	
tip3p.itp	tip3p.itp	
ions.itp	ions.itp	

File domain.ndx uploaded.

Files successfully uploaded

Here is a brief overview of your files:

The topology lists 109949 molecules:

Protein_chain_A: 1

SOL: 109923

NA: 25

The index file lists 20 groups

FIGURE 1 Short system summary after completed file upload for the DnaK test case.

Network setup

After completing the file upload, the user can set up the network. The standard approach is to construct the network at the residue level, with each network node representing one residue, as specified in the user-provided topology. The user may also employ custom node definitions, as outlined below.

The GROMACS index file (.ndx) uploaded by the user is used to define the network. This file lists groups of atoms, such as all protein atoms or all atoms belonging to a given domain. The user can create custom groups with the GROMACS tool `make_ndx`. MDN reads the index file and checks groups that can be used to create a network, with the requirement that the group contains all the atoms of a given residue if any of these atoms is listed. A typical group choice for network construction is shown in Fig. 2 A, in which the entire protein for the DnaK test case is used to create the network, thus generating 600 nodes. The user is also able to construct smaller networks, including only residues belonging to a specific domain. This is shown in Fig. 2 B, where a subset of protein residues (the group

A

Index	Group	Selection
1	Protein	<input checked="" type="checkbox"/>
10	Prot-Masses	<input type="checkbox"/>
14	non-Water	<input type="checkbox"/>
15	Ion	<input type="checkbox"/>
16	NA	<input type="checkbox"/>
18	SBD	<input type="checkbox"/>
19	NBD	<input type="checkbox"/>

B

Index	Group	Selection
1	Protein	<input type="checkbox"/>
10	Prot-Masses	<input type="checkbox"/>
14	non-Water	<input type="checkbox"/>
15	Ion	<input type="checkbox"/>
16	NA	<input type="checkbox"/>
18	SBD	<input checked="" type="checkbox"/>
19	NBD	<input type="checkbox"/>

FIGURE 2 Network setup for the DnaK test case. (A) Network construction for group Protein (600 nodes). (B) Network construction for group SBD (108 nodes).

representing the substrate-binding domain (SBD) of DnaK) is used to construct the network, thus generating 108 nodes. We note that choosing a specific group will result in a network constructed by splitting the group over its constituent residues. The user may also choose to upload an index file with custom node definitions, in which case the only group available for selection will be named SPECIFIED_NODES. Instructions for using custom node definitions can be found in the MDN documentation.

We note that calculating the interaction energies from the uploaded trajectories is the most demanding step of the entire protocol. Depending on the number of trajectory frames and the complexity of the molecular system and its interactions, this step may take 5 min or longer. The user is able to follow the overall progress.

Choosing pathways

The analysis of network properties depends on the network topology and the subset of possible pathways that is analyzed. A network containing N nodes has $N(N - 1)/2$ pairs of nodes, each giving rise to a pathway. MDN lets the user choose whether all these pathways should be calculated and analyzed. This is again accomplished through the index file, and a group is allowed if it is a subset of the group used for network construction. It should be noted that the same input index (.ndx) file used to define the nodes (see previous section) will be used here to define the groups of nodes that will be considered for calculation of the pathways. Therefore, the user should decide on these groups at the moment he/she is making the index file.

A typical choice for pathway calculation is shown in Fig. 3 A, where all pathways for the network containing the entire protein are calculated (179,700 pathways). Fig. 3 B is an example where two nonoverlapping regions of the network are selected for analysis. In this case, pathways between groups SBD (108 nodes) and the nucleotide-binding domain (NBD) (380 nodes) are calculated, yielding a total of 74,860 pathways.

Output

Once all the information described above is specified, the program performs the energy network analysis and the user can download a zip file containing several analysis output files. A text file containing node betweenness and normalized node betweenness η values for each network node is given. The same information is also given in a file listing every atom belonging to residues represented in the network. A Python script, which is also included, can be used to generate a PDB file containing this information in the occupancy field. This allows the user visualization of the node betweenness using any of the biomolecular visualization programs available. Information about the global network coupling, as well as partial network coupling,

A			
Index	Group	Selection 1	Selection 2
1	Protein	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
10	Prot-Masses	<input type="checkbox"/>	<input type="checkbox"/>
18	SBD	<input type="checkbox"/>	<input type="checkbox"/>
19	NBD	<input type="checkbox"/>	<input type="checkbox"/>

B			
Index	Group	Selection 1	Selection 2
1	Protein	<input type="checkbox"/>	<input type="checkbox"/>
10	Prot-Masses	<input type="checkbox"/>	<input type="checkbox"/>
18	SBD	<input checked="" type="checkbox"/>	<input type="checkbox"/>
19	NBD	<input type="checkbox"/>	<input checked="" type="checkbox"/>

FIGURE 3 Selecting nodes for pathway calculation for the DnaK test case. A) Example showing group choice that leads to all pathways for the group called Protein being calculated. B) Example showing choice that leads to pathways between two non-overlapping groups (SBD and NBD) being calculated.

values between different groups listed in the index file is given in another file. Finally, files containing information about the coupling between each node of the network and the groups listed in the index file are provided. Each of these files lists every atom belonging to residues represented in the network, along with the coupling value of the corresponding node. These files can be used to generate maps showing the energetic coupling to a specific region over the entire structure of the molecular system.

In the next sections, we outline the usage of MDN to analyze molecular dynamics simulations of the WT and three mutants of DnaK.

Application to analysis of Hsp70 variants

The Hsp70 chaperones are essential for protein transport and folding and have been shown to be the most conserved proteins in all organisms (34,35). They are essential for cell survival, as they help cells withstand stresses by helping unfolded/misfolded proteins fold properly. For this reason, inhibition of Hsp70 activity has been suggested as a treatment for cancer (36) and Alzheimer's disease (37). The DnaK structure, shown in Fig. 4, has three domains: the NBD, the SBD, and an α -helical C-terminal domain. This is an important example of an allosteric system, as ATP binding at the NBD triggers substrate release at the SBD, thus involving communication over a distance of 50 Å. In addition, motivation to understand allosteric signaling in Hsp70 comes from the idea that in designing inhibitors of Hsp70 function, allosteric drugs might offer higher selectivity than drugs designed to bind to the molecule's primary binding sites (34).

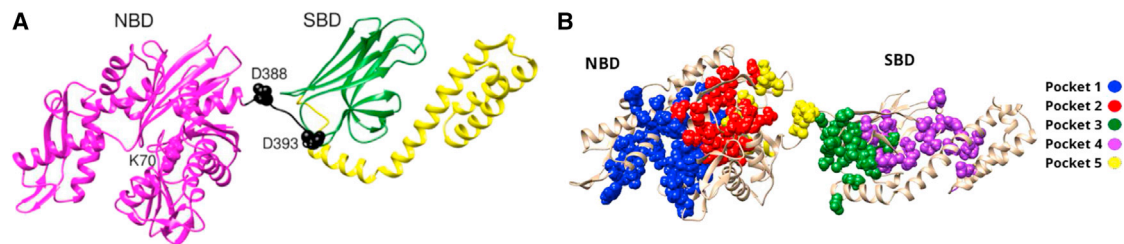


FIGURE 4 DnaK structure (PDB 2KHO). (A) Visualization of the three domains: the NBD (*magenta*), the SBD (*green*), and the C-terminal domain (*yellow*). The NBD-SBD linker is colored in black. Residues K70, D388, and D393 are shown as spheres. (B) Visualization of the five pockets identified by Rodina and co-workers (40). Residues lining the pockets are represented as spheres colored according to the color code. To see this figure in color, go online.

Experimental attempts to characterize the sources of Hsp70 activity have shown that certain single-point mutations can affect both ATP binding and interdomain communication, leading to inactive Hsp70 mutants. We have performed MD simulations of four experimentally studied DnaK variants: WT, D388R and D393R (38), and K70E (39). The WT and D388R variants were shown to be allosterically active, whereas the remaining variants are inactive. It is particularly interesting to note that the mutation from Asp to Arg at position 388 does not suppress allosteric activity, whereas the opposite is true at position 393. Both residues are part of the linker that connects the NBD with the SBD (Fig. 4, *black*).

MDN was used to calculate the partial network energetic coupling between each amino acid residue in the NBD and the entire SBD. Fig. 5 shows the results, with each residue in the NBD colored according to the corresponding coupling value. It can be seen that the coupling values are higher for the two active proteins, WT and D388R.

A recent publication by Rodina and co-workers (40) reported on a potential Hsp70 inhibitor found by designing a drug that would fit to a cavity found in the protein by using software called SiteMap. This software finds cavities in the molecule and uses local information (size, hydrophobicity, etc.) to determine their druggability, a measure of the probability that the drug will bind to a cavity (41). However, because it does not consider the global response of the protein to the binding of a drug to a particular cavity, the software cannot make predictions on the probability that the drug will exert the desired effect on the molecule (which, in the case of Hsp70, is to block the allosteric communication between the domains). In the work of Rodina and co-workers, SiteMap was used to identify the five most druggable pockets in Hsp70 (Fig. 4 B), and the authors then picked one pocket based on considerations of size and hydrophobicity and performed experiments to check that the drug binds to Hsp70 and to test for Hsp70 inhibition.

We calculated the energetic coupling between the residues that line each of the five pockets (P.D. Pattel, Memorial Sloan-Kettering Cancer Center and St. John's University, personal communication, 2014) and the other domain of

Hsp70 (i.e., if the pocket was in the NBD, we would calculate its coupling with the SBD, and vice versa). Table 1 shows the values obtained. Pockets 1 and 4 show the best communication with the other domain. Of these, pocket 4 is a small cavity, whereas pocket 1, chosen by the investigators, shows good communication with the SBD domain (0.64) and is a large cavity, so we would also have chosen pocket 1 for the target site. This suggests that energetic coupling could offer some predictive power over the response the molecule will have to binding of a small molecule to one of its cavities, before going into the experimental validation phase.

CONCLUSIONS

We have built an online tool, called MDN (<http://mdn.cheme.columbia.edu>), that allows the application of energy-based network theory to the analysis of molecular dynamics trajectories of proteins, for the characterization of allosteric signaling pathways. The tool employs a method recently published by our group (21), which assigns network edge weights based on the strength of energetic interaction between residues (used to define the network nodes). MDN provides an interface that allows the user to upload all the necessary data, as well as define the groups of residues that are to be considered in the analysis. Its output includes values for the node betweenness (21) (a measure of the importance of a residue for efficient signal propagation in a protein), and the energetic coupling (22) (a measure of efficiency of signal propagation in a protein network). MDN can be used to analyze output from simulations performed with GROMACS and NAMD, and future improvements will include support for additional simulation packages.

We used MDN to characterize allosteric signaling in the WT and three mutants of DnaK. Our calculations of energetic coupling between the NBD and the SBD in DnaK correlate well with experimental results. Two allosterically active mutants showed high values for the energetic coupling, whereas two allosterically inactive mutants showed lower values of this quantity. In addition, a pocket recently used to design an allosteric inhibitor for Hsp70

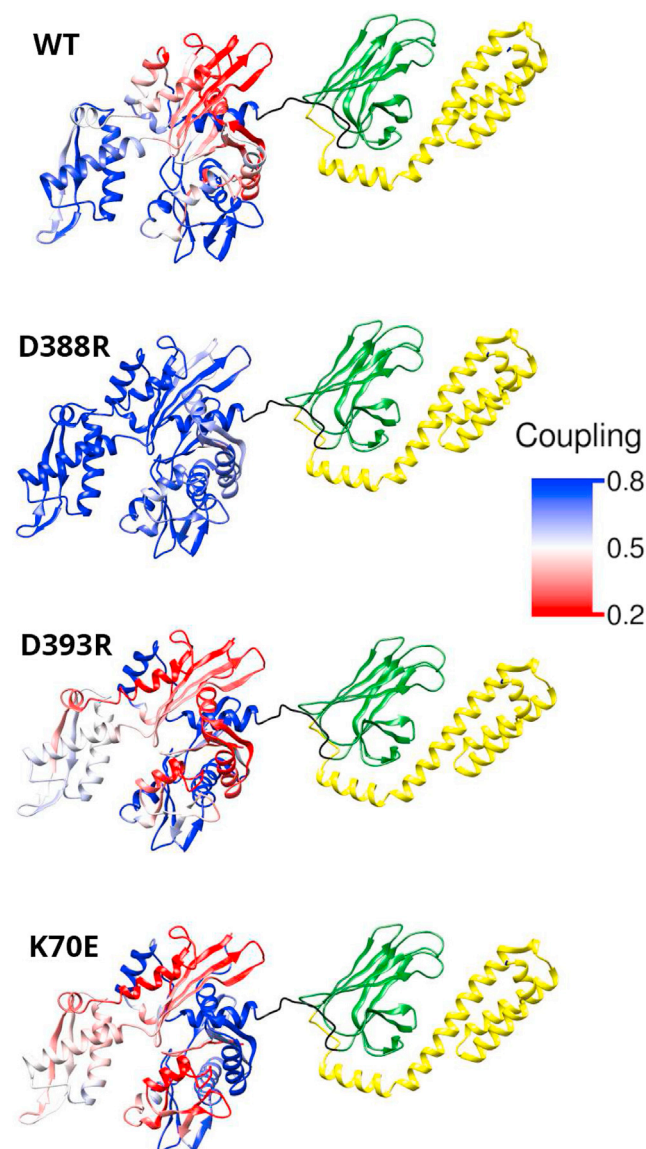


FIGURE 5 Maps of energetic coupling for DnaK WT and for mutants experimentally found to be allosterically active (D388R) or inactive (D393R and K70E). Each amino acid in the NBD is colored according to the energetic coupling calculated between that amino acid and all amino acids forming the SBD (green). To see this figure in color, go online.

also showed high values of energetic coupling. This suggests that energetic coupling could be used as a measure of druggability of a pocket to predict the response of the protein to drug binding before going into the (more expensive) experimental validation phase.

TABLE 1 Partial energetic couplings between residues surrounding each pocket and the other domain of Hsp70

Pocket	1	2	3	4	5
Domain	SBD	SBD	NBD	NBD	SBD
Coupling	0.64 ±0.05	0.42 ±0.06	0.43 ±0.08	0.66 ±0.05	0.36 ±0.06

AUTHOR CONTRIBUTIONS

A.A.S.T.R. and V.O. designed the research; A.A.S.T.R. designed and implemented the MDN server tool, performed the simulations, and analyzed the data; A.A.S.T.R. and V.O. wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank Dr. Gabriela Chiosis and Dr. Pallav Patel for kindly providing information about the residues lining the five Hsp70 pockets presented in their report.

This material is based upon work supported by a Parkinson's Disease Foundation Columbia University Center Pilot grant program.

REFERENCES

- Lindahl, E. 2015. Molecular dynamics simulations. *In* *Molecular Modeling of Proteins*. Springer, New York, pp. 3–26.
- Bernardi, R. C., M. C. R. Melo, and K. Schulten. 2014. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta.* 1850:872–877.
- Case, D. A., T. E. Cheatham, 3rd, ..., R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
- Brooks, B. R., C. L. Brooks, 3rd, ..., M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.
- Pronk, S., S. Páll, ..., E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 29:845–854.
- Kunz, A.-P. E., J. R. Allison, ..., W. F. van Gunsteren. 2012. New functionalities in the GROMOS biomolecular simulation software. *J. Comput. Chem.* 33:340–353.
- Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
- Balsera, M. a., W. Wriggers, ..., K. Schulten. 1996. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100:2567–2572.
- Collier, G., and V. Ortiz. 2013. Emerging computational approaches for the study of protein allostery. *Arch. Biochem. Biophys.* 538:6–15.
- Ota, N., and D. A. Agard. 2005. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* 351:345–354.
- Sharp, K., and J. J. Skinner. 2006. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins.* 65:347–361.
- Ming, D., and M. E. Wall. 2005. Quantifying allosteric effects in proteins. *Proteins.* 59:697–707.
- Chennubhotla, C., and I. Bahar. 2006. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Sys. Biol.* 2:36.
- Chennubhotla, C., Z. Yang, and I. Bahar. 2008. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosyst.* 4:287–292.
- Ribeiro, A. A. S. T., and V. Ortiz. 2015. Local elastic constants of LacI and implications for allostery. *J. Mol. Graph. Model.* 57:106–113.
- Sethi, A., J. Eargle, ..., Z. Luthey-Schulten. 2009. Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. USA.* 106:6620–6625.
- Gaspar, P. M., B. Fuglestad, ..., J. A. McCammon. 2012. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc. Natl. Acad. Sci. USA.* 109:21216–21222.
- Freddolino, P. L., K. H. Gardner, and K. Schulten. 2013. Signaling mechanisms of LOV domains: new insights from molecular dynamics studies. *Photochem. Photobiol. Sci.* 12:1158–1170.

19. Vanwart, A. T., J. Eargle, ..., R. E. Amaro. 2012. Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Comput.* 8:2949–2961.
20. Ghosh, A., and S. Vishveshwara. 2007. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA.* 104:15711–15716.
21. Ribeiro, A. A. S. T., and V. Ortiz. 2014. Determination of signaling pathways in proteins through network theory: importance of the topology. *J. Chem. Theory Comput.* 10:1762–1769.
22. Ribeiro, A. A. S. T., and V. Ortiz. 2015. Energy propagation and network energetic coupling in proteins. *J. Phys. Chem. B.* 119:1835–1846.
23. Bertelsen, E. B., L. Chang, ..., E. R. P. Zuiderweg. 2009. Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl. Acad. Sci. USA.* 106:8471–8476.
24. Pettersen, E. F., T. D. Goddard, ..., T. E. Ferrin. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
25. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
26. Duan, Y., C. Wu, ..., P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.
27. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
28. Parrinello, M., and A. Rahman. 1980. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* 45:1196–1199.
29. Hess, B. 2008. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122.
30. Miyamoto, S., and P. A. Kollman. 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.
31. Boccaletti, S., V. Latora, ..., D. Hwang. 2006. Complex networks: structure and dynamics. *Phys. Rep.* 424:175–308.
32. Floyd, R. W. 1962. Algorithm 97: shortest path. *Commun. ACM.* 5:345.
33. Latora, V., and M. Marchiori. 2001. Efficient behavior of small-world networks. *Phys. Rev. Lett.* 87:198701.
34. Zuiderweg, E. R. P., E. B. Bertelsen, ..., A. Ahmad. 2013. Allostery in the Hsp70 chaperone proteins. *Top. Curr. Chem.* 328:99–153.
35. Gupta, R. S. 1998. Protein phylogenies and signature sequences: a re-appraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62:1435–1491.
36. Murphy, M. E. 2013. The HSP70 family and cancer. *Carcinogenesis.* 34:1181–1188.
37. Jinwal, U. K., Y. Miyata, ..., C. A. Dickey. 2009. Chemical manipulation of hsp70 ATPase activity regulates τ stability. *J. Neurosci.* 29:12079–12088.
38. Vogel, M., M. P. Mayer, and B. Bukau. 2006. Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. *J. Biol. Chem.* 281:38705–38711.
39. O'Brien, M. C., K. M. Flaherty, and D. B. McKay. 1996. Lysine 71 of the chaperone protein Hsc70 is essential for ATP hydrolysis. *J. Biol. Chem.* 271:15874–15878.
40. Rodina, A., P. D. Patel, ..., G. Chiosis. 2013. Identification of an allosteric pocket on human hsp70 reveals a mode of inhibition of this therapeutically important protein. *Chem. Biol.* 20:1469–1480.
41. Halgren, T. A. 2009. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* 49:377–389.