Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Probing allosteric communication with combined molecular dynamics simulations and network analysis

Mattia Bernetti[1,2], Stefano Bosio[1,2], Veronica Bresciani[1,2], Federico Falchi[1,2] and Matteo Masetti[1]

Abstract

Understanding the allosteric mechanisms within biomolecules involved in diseases is of paramount importance for drug discovery. Indeed, characterizing communication pathways and critical hotspots in signal transduction can guide a rational approach to leverage allosteric modulation for therapeutic purposes. While the atomistic signatures of allosteric processes are difficult to determine experimentally, computational methods can be a remarkable resource. Network analysis built on Molecular Dynamics simulation data is particularly suited in this respect and is gradually becoming of routine use. Herein, we collect the recent literature in the field, discussing different aspects and available options for network construction and analysis. We further highlight interesting refinements and extensions, eventually providing our perspective on this topic.

Addresses

[1] Department of Pharmacy and Biotechnology, Alma Mater Studiorum - University of Bologna, Via Belmeloro 6, 40126 Bologna, Italy

[2] Computational and Chemical Biology, Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

Corresponding authors: Masetti, Matteo (matteo.masetti4@unibo.it); Bernetti, Mattia (mattia.bernetti2@unibo.it)

(Bernetti M.), (Bosio S.), (Bresciani V.)

Current Opinion in Structural Biology 2024, **86**:102820

This review comes from a themed issue on **New Concepts in Drug Discovery (2024)**

Edited by **Andrea Cavalli** and **Alessio Ciulli**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2024.102820

Introduction

The word allosteric was coined by Monod and Jacobs to describe a mechanistic model previously proposed by Changeux for "apparently competitive" enzyme inhibition [1]. The term was meant to convey the presence of two topologically distinct binding sites, one for the substrate and one for the inhibitor (more in general, the allosteric effector), and that the inhibition was mediated by a conformational change of the enzyme, rather than by mutual exclusion on the catalytic site [1]. The model was then generalized envisioning an equilibrium of two pre-existing protein conformational states with different affinities for both the substrate and effector, and that the preferred conformation is *selected* upon effector binding [2]. Such a model, known as MWC (Monod-Wyman-Changeux), is in sharp contrast with the almost contemporary KNF (Koshland-Némethy-Filmer) model, in which the allosteric conformational transition is rather seen as *induced* by the effector [3]. Over the years, the concept of allostery has significantly evolved, including not only conformational changes but also changes in conformational entropy (usually referred to as entropically-driven allostery or "dynamic allostery") [4,5]. Furthermore, it has been proposed that all proteins are potentially allosteric [6], underscoring the importance of allosteric communication in their mechanistic functioning [7].

Modulating the function of proteins involved in diseases is ultimately the goal of drug discovery, and indeed, several drugs bind to allosteric effector sites, exploiting allostery as their mechanism of action [8]. Allosteric drugs offer several potential advantages over conventional drugs [9]. For example, since allosteric sites are expected to be less evolutionarily conserved than enzyme active sites or receptor orthosteric sites, allosteric drugs hold great potential whenever selectivity toward a specific member of a protein family is required. Furthermore, allosteric drugs can be used to circumvent drug resistance to orthosteric sites [8]. From a computational standpoint, the identification of allosteric pockets is as important as the identification of potential binders at those sites. However, while the latter can be carried out with well-established methodologies, like molecular docking and virtual screening, the identification of allosteric sites is less consolidated.

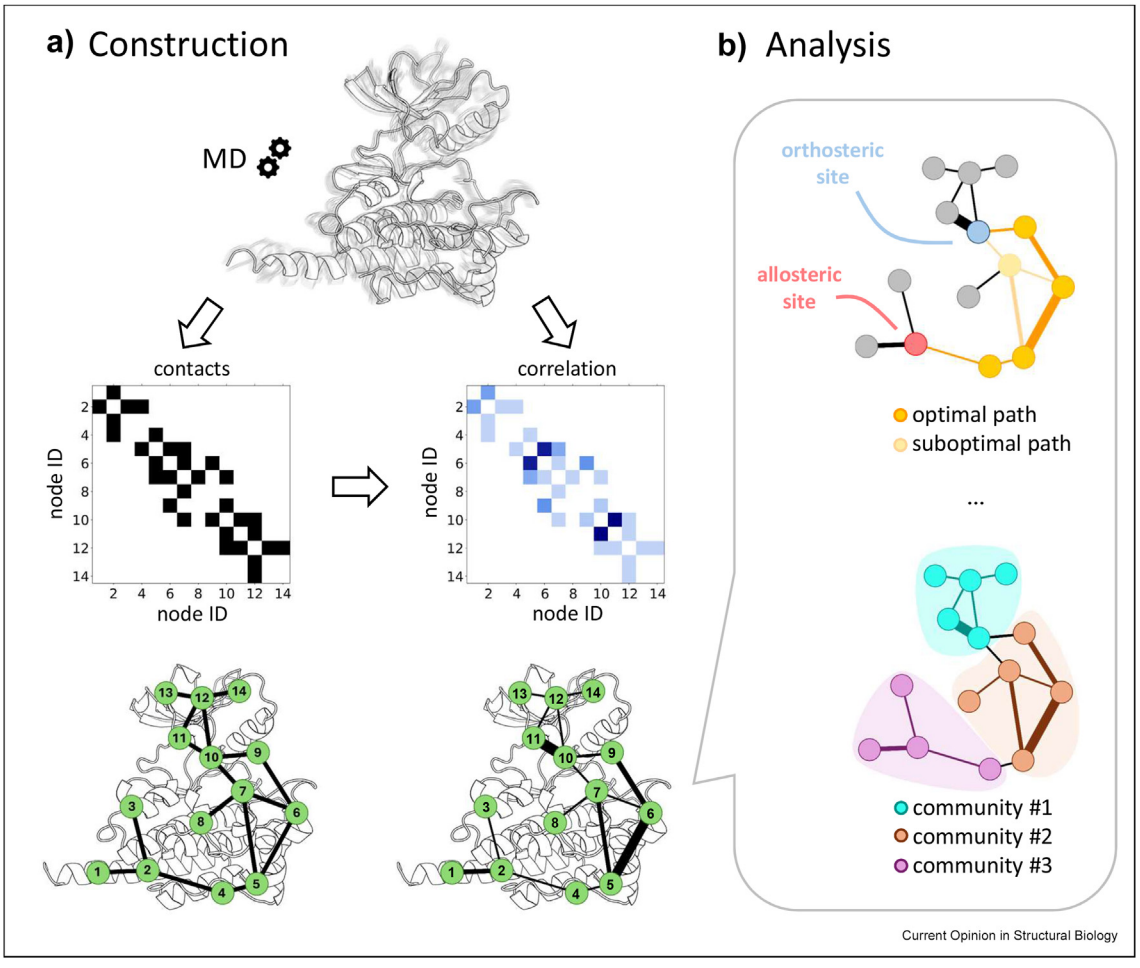
The identification of allosteric sites is a twofold problem: pocket identification and characterization of the allosteric communication involved by targeting that pocket [10]. Here we focus on a subset of the several computational methods available for covering the latter aspect [11]. Considering the intimate relationship

between allostery and protein dynamics, Molecular Dynamics (MD) simulations have emerged as a natural choice to probe allosteric communication. Thus, several computational methods based on MD sampling, but relying on different conceptual frameworks, have been developed over the years for disentangling allostery [12]. Recognizing that proteins can be viewed as networks of interacting nodes, herein we focus on the methods combining MD simulations and network analysis to investigate allosteric communication, and review their recent applications to pharmaceutically relevant systems. Following the recent surge of artificial intelligence (AI) to tackle complex problems in the life sciences, approaches leveraging AI are also increasingly being applied to investigate allostery. For further insights into this emerging class of methods, that are not covered in this review, we remand to recent exhaustive surveys [13,14].

Building and analyzing a correlation-based network

The underlying assumption behind the computational methods covered in this review is that the communication between allosteric and functional sites can be detected through MD simulations even without explicitly observing the entire allosteric transition, which is generally expected to occur in timescales that are difficult to sample. The study of allostery therefore focuses on the identification of coupled motions between topologically distinct structural elements from the analysis of inter-residue correlation data stored in

Figure 1



General workflow to perform network analysis on MD simulation data. (**a**) In the network construction stage, MD trajectories (top structure) are used to build i) a contact map (black and white matrix, on the left) that defines edges connecting nodes and ii) a filtered correlation matrix (in blue color scale, on the right) with edge weights. The nodes (green circles with labeled node index, on the bottom structures) represent residues of the biomolecule. (**b**) Different analyses can be then performed on the constructed network to extract relevant information, such as the study of optimal and suboptimal pathways or the identification of node communities. Here, we used a toy network for illustrative purposes. As shown in the schematic, optimal pathways (dark orange) in weighted networks may encompass a higher number of edges compared to suboptimal ones (yellow), as long as information transfer along those edges is more efficient. Different options exist to partition the network into communities, and in this example we used the popular Girvan–Newman algorithm.

the MD trajectory (Figure 1(**a**)) [7]. Such correlated motions are then framed in terms of a network that is further interrogated through conventional graph theory algorithms or *ad hoc* tools (Figure 1(**b**)). Additionally, the effect of ligands or mutations on the allosteric communication can be rationalized by comparing networks derived from simulations carried out under different conditions (e.g. unbound protein, effector-bound protein, etc.).

The network is built by mapping the N residues of a protein into N nodes of a weighted graph (Figure 1(**a**), green circles with labeled node index on the bottom structures). The information regarding connectivity between nodes is specified in the adjacency matrix, a square symmetric $N \times N$ matrix, where, in its simplest form, the elements represent the pairwise contacts established during the simulation (Figure 1(**a**), black and white matrix and respective bottom network). Such a contact map matrix requires three adjustable parameters: *i*) the distance cutoff used to consider a contact as formed in a given instant of time, *ii*) the atoms among which such a distance cutoff must be evaluated, and *iii*) the fraction of simulation time that is used to consider a contact as statistically relevant [15]. Typically, the contacts are evaluated by measuring the distance between Cα atoms of non-consecutive residues, but the residue center of mass as well as the minimum distance among all the heavy atoms of the considered residues can also be used (see Table 1). We note that the residue atoms (or groups thereof) employed for constructing the contact map are not necessarily the same ones used in the graphical representation of the nodes of the network, for which the Cα atoms are mostly employed.

The edges between nodes, which are binary in the contact map, are then weighted to return a continuous range of values representing the probability of information transfer encoded by the degree of correlated motions (Figure 1(**a**), matrix in blue color scale and respective bottom network). In this respect, the contact map serves as a filter for the correlations based on their physical proximity. Typically, the edge weight is set as: $w_{ij} = -ln(|c_{ij}|)$, where c_{ij} is the normalized covariance (or cross-correlation) [15]. However, different metrics can be adopted for both the edge weight and the measure of correlation. For example, the generalized correlation coefficient [67], which is based on the concept of mutual information (MI) [68], allows the detection of both linear (rLMI) and non-linear (rMI) relationships, and is preferred over cross-correlation in most recent applications (see Table 1) [16–19,28–30,38–43,46,55,64,69]. Furthermore, the bare MI can be directly used [21,22,27,47,49–51,63,65,66,70,71]. In any case, we stress that these correlation metrics are based on equilibrium properties of individual conformational ensembles, rather than dynamic quantities [5].

Allosteric communication can be finally characterized by identifying communication pathways, critical nodes, or by partitioning the network into communities (Figure 1(**b**)). Communication pathways are determined between pairs of nodes, usually termed source and target (or sink), which are expected to act as endpoints for the allosteric communication (Figure 1(**b**), top schematics). The Floyd-Warshall [16,18,19,29,38–41,43–45,53] and the Dijkstra [25,28,37,46,55,61,65,66] algorithms are well-established choices to identify optimal (i.e. shortest) pathways in the network. Inspecting a number of suboptimal pathways, which are closest in length to the optimal one, can also be crucial to pinpoint alternative routes for communication transfer or to highlight recurring nodes. Available options for this scope are the Yen's K-shortest path [48,52], the Weighted Implementation of Suboptimal Paths (WISP) [23,24,53,62,72], and the Subset of Adjacent Nodes (SOAN) algorithms [46,73]. Relevant nodes in information transfer can also be identified via centrality measures. The betweenness centrality of a node, in particular, quantifies the number of unique shortest paths passing through that node [31,37,47,61,64–66].

Differently, community analysis provides a coarse-grained picture of the network, where highly connected nodes are grouped together into distinct communities (Figure 1(**b**), bottom schematics). This can prove particularly beneficial in disentangling the information transfer between regions of large biomolecular assemblies, especially when limited knowledge about the system hinders the focus on specific nodes. This task typically leverages the Girvan-Newman algorithm [25,31,47,48,53], which uses as partitioning criterion the edge betweenness (EB), i.e. the number of shortest paths crossing a given edge. The resulting community network can provide a simplified representation, where nodes correspond to different communities and edges are weighted by an inter-community EB. An interesting alternative is the more heuristic Louvain algorithm [29,32], based on modularity optimization, demonstrating an improved performance that can be suitable to larger macromolecular complexes comprising large amounts of nodes. Lastly, communities can be intentionally enforced to observe communication exchange between specific regions, such as domains or binding pockets [38,39,42].

As a final note, we wish to remark that the results that can be obtained by such a framework composed of network construction and analysis are ultimately dependent on the underlying MD sampling. The practice of performing MD simulation replicates, which is always advisable to mitigate sampling limitations, can be beneficial also in this context. Additionally, a more thorough exploration of the conformational space can be achieved through

Table 1

Recent applications of network analysis based on MD simulations to study allosteric mechanisms in pharmaceutically relevant systems. For each of the reported articles, the table reports information on network construction (features, metrics, and contact map type), on the underlying MD sampling, on the presence of ligands (in the simulations, and possibly in the network as nodes), on the type of path and community analyses, on the application of multi-ensemble strategies, and on the use of specific software. Abbreviations: i) continuous (cont.), discontinuous (disc.), ii) Mutual Information (MI, rMI if generalized), Linear Mutual Information (LMI, rLMI if generalized), Cross-correlation (c), Distance Fluctuations (DF), Local Spatial Patterns (LSP); iii) Hamiltonian Replica Exchange (HREX), metadynamics (metad), accelerated MD (aMD), Gaussian accelerated MD (GaMD); iv) Floyd-Warshall (FW), Dijkstra (DK), Weighted Implementation of Suboptimal Paths (WISP), Yen's K-Shortest Path (YK), Betweenness Centrality (BC), Subset of Adjacent Nodes (SOAN), Signal-to-Noise Ratio (SNR); v) Girvan-Newman (GN), Louvain (LV), Guimerà-Amaral Cartography (GA), Hierarchical Clustering (HC); vi) Dynamical Perturbation Contact Network (DPCN), difference Contact Network Analysis (dCNA). * indicates ligands different from small molecules, such as ions, lipids, sugars, peptides.

Publication	Features	Contact map	Metric	MD sampling	Has ligands	Ligand as nodes	Path analysis	Community analysis	Multi-ensemble	Software
Naseem-Khan [16]	Cα, P, N1	disc.	rMI	Plain	Yes	No	Optimal (FW)	–	–	DyNetAn
Ginex [17]	Cα	disc.	rMI	Plain	No	–	–	–	–	DyNetAn
Tajima [18]	Cα	disc.	rMI	Plain	No	–	Optimal (FW)	–	–	DyNetAn, NetworkX
Krishnan [19]	Cα	disc.	rMI	Plain	No	–	Optimal (FW)	–	–	RING, NetworkX
Liu [20]	Dihedrals	cont.	1/c	Plain	Yes	No	–	–	–	–
Kinnebrew [21]	Cα	cont.	MI	Plain	Yes	Yes	Information flow	–	–	Allopath
Zhang [22]	Cα	disc.	c, MI	Plain	Yes	Yes	Optimal (FW), Suboptimal (FW), information flow	–	–	Network view
Kihn [23]	Residue COM	disc.	c	Ratcheted MD	Yes	No	Suboptimal (WISP)	–	–	WISP
Sun [24]	Cα	disc.	c	GaMD	Yes*	No	Suboptimal (WISP)	–	–	WISP
Calvó-Tusell [25]	Cα	disc.	c	aMD	Yes	No	Optimal (DK)	GN	–	SPM Web Tool
Gheeraert [26]	Heavy atoms	disc.	no. contacts	Plain	Yes*	No	–	–	DPCN	–
Kornev [27]	Cartesian and internal	disc.	LSP, LMI, c	Plain	Yes	Yes	–	GA	–	Bio3D, rnetcarto
Maschietto [28]	Cα	disc.	rMI	Plain	Yes	No	Optimal (DK)	–	–	–
Yang [29]	Cα	disc.	rLMI	Plain	Yes*	Yes	Optimal (FW), Suboptimal (YK)	LV	–	Correlationplus, Gephi, NetworkX
Ray [30]	Cα	disc.	rLMI	Plain	Yes	No	–	–	–	Bio3D
Santos [31]	Heavy atoms	disc.	c	Plain	Yes	No	BC	GN	–	Bio3D
Barbera [32]	Beads	disc.	no. contacts	Plain	Yes	No	–	LV	dCNA	adapted from dCNA
Fung [33]	Cα	disc.	no. contacts	Plain	Yes*	No	–	GN	dCNA	Bio3D, dCNA
Li [34]	Heavy atoms	disc.	no. contacts	GaMD/Plain	Yes	No	Suboptimal (YK)	GN	dCNA	dCNA
Kumutima [35]	Heavy atoms	disc.	no. contacts	Plain	Yes*	No	Suboptimal (YK)	–	dCNA	Bio3D
Yao [36]	Heavy atoms	disc.	no. contacts	Plain	Yes	Yes	Suboptimal (YK)	–	dCNA	dCNA
Costa [37]	Heavy atoms	disc.	c	Plain	No	–	Optimal (DK), BC	–	–	NetworkX
Sinha [38]	Cα, P, N1, N9	disc.	rMI	Plain	No	–	Optimal (FW), Suboptimal (YK), SNR	Enforced, for SNR	–	DyNetAn, NetworkX
Molina Vargas [39]	Cα, P, N1, N9	disc.	rMI	Plain	No	–	Optimal (FW), Suboptimal (YK), SNR	Enforced, for SNR	–	DyNetAn, NetworkX
Liu [40]	Cα, P, N1, N9	disc.	rMI	Plain	No	–	Optimal (FW)	–	–	DyNetAn
Yovanno [41]	Cα	disc.	rMI	Metad	No	–	Optimal (FW)	–	–	DyNetAn

Aguti [42]	Heavy atoms	disc.	rMI, DF, pocket cross-talk	Plain	No	–	–	Enforced as pocket residues	–	DyNetAn, Pocketron, NetworkX
Li [43]	Cα	disc.	rMI	GaMD	Yes	Yes	Optimal (FW)	–	–	DyNetAn
Spinello [44]	Residue COM	disc.	c	Plain	No	–	Optimal (FW), Suboptimal (FW)	–	–	CPPtraj, DyNetAn
Dube [45]	Heavy atoms	disc.	c	Plain	Yes	Yes	Optimal (FW)	–	–	CARMA, Network View
Konovalov [46]	Cα and one side chain heavy atom	cont.	rLMI	Plain	No	–	Optimal (DK), Suboptimal (SOAN), Path lumping	–	–	g_correlation
Soya [47]	Cα	disc.	LMI	Plain	Yes	No	BC	GN	–	wordom, Bio3D
Dayananda [48]	Cα	disc.	c	Plain	No	–	Suboptimal (YK)	GN	–	Bio3D
Janaszkiewicz [49]	Residue COM	cont.	MI	Plain	Yes	Yes	Information flow	–	–	Allopath
Tóth [50]	Residue COM	cont.	MI	Plain	Yes	Yes	Information flow	–	–	Allopath
Tóth [51]	Residue COM	cont.	MI	Plain	Yes*	Yes	Information flow	–	–	Allopath
Cheng [52]	Backbone beads	disc.	c	Plain	No	–	Suboptimal (YK)	HC	–	Bio3D
Omotuyi [53]	Cα	disc.	c	Adaptive sampling/Plain	Yes	No	Optimal (FW), Suboptimal (WISP)	GN	–	Bio3D, WISP
Gheeraert [54]	Heavy atoms	disc.	no. contacts	Plain (T = 303.15 K, 323.15 K)	No	–	–	Connected Component Analysis	DPCN	Scipy, MDtraj, NetworkX
Maschietto [55]	Cα; Heavy atoms	disc.	rMI; no. contacts	Plain (T = 303.15 K, 323.15 K)	Yes	No	Optimal, Suboptimal (DK)	–	DPCN	–
Chen [56]	Cα	disc.	no. contacts	Plain	No	–	–	GN	dCNA	dCNA, Bio3D
Ouedraogo [57]	Heavy atoms	disc.	no. contacts	Plain	Yes	No	–	GN + network modularity	dCNA	dCNA, igraph, Bio3d
Pegram [58]	Heavy atoms	disc.	no. contacts	Plain (T = 285 K, 300 K, 315 K, 330 K)	No	–	–	–	dCNA	dCNA, MDAnalysis
Souffrant [59]	Heavy atoms	disc.	no. contacts	Plain	No	–	–	GN + network modularity	dCNA	dCNA, igraph, Bio3d
Yu [60]	Cα, P	disc.	no. contacts	Plain	No	–	–	GN	dCNA	MDTraj
Kelly [61]	Cα; side chain heavy atoms	disc.	c; no. contacts	Plain	Yes	No	Optimal (DK), Suboptimal (DK), BC	–	–	NetworkX
Crean [62]	Heavy atoms	disc.	c	HREX	Yes*	No	Suboptimal (WISP)	–	–	MDTraj, pycontact, Bio3D
Srivastava [63]	Cα, P	cont.	LMI	Plain	No	–	–	–	–	–
Jani [64]	Heavy atoms	disc.	rMI	Plain	Yes	No	BC	–	–	Bio3D
Bassetto Jr [65]	Side chain COM, Cα	cont.	MI	Plain	No	–	Optimal (DK), BC	–	–	NetworkX
Costa [66]	Side chain COM, Cα	cont.	MI	Plain	Yes	Yes	Optimal (DK), BC	–	–	NetworkX

enhanced sampling approaches, when needed [23–25,34,40,41,43,53,62].

Network refinements and extensions

Herein, we briefly summarize noteworthy modifications or additions that have been introduced over the years at different levels of the above-reported general scheme on network construction and analysis:

Refinements/extensions in network construction:

- The contact map defined via the distance and statistical time-percentage cutoffs results in a sparse network, where each node is not necessarily connected with all the other ones. In an effort to improve the robustness of the contact map, and in turn of the whole network topology, continuous alternatives, leveraging different smoothing functions, have been proposed [70,71,74] and used [20,21,46,49–51,63,65,66].
- In most cases, the correlation metric is applied to the displacement of Cartesian coordinates. Nevertheless, different features can be effectively employed, such as dihedral angles or electrostatic energies between residue pairs. Interestingly, calculation of all these features is implemented in the recent MDiGest package [75].
- Nodes typically represent residues of the biomolecule. Recently, ligands or structural ions have also been included explicitly in the network as nodes [21,22,27,29,36,43,45,49–51,66,70]. This can be convenient to inspect pathways of communication associated with ligand binding, but the standard approach may be preferred to compare straightforwardly ligand-bound and unbound ensembles [67]. In the specific case of membrane systems, such as GPCRs or ion channels and transporters, lipids are emerging as important players of allosteric modulation. When they are mapped as nodes of the network, permutation invariance must be taken into account to cope with their possible exchange during the simulation [70].

Refinements/extensions in the analysis of the networks:

- Alternative methods have recently emerged for the identification of critical nodes. Eigenvector centrality based on mutual information, achieved through diagonalization of the adjacency matrix, measures how well connected a node is to other well-connected nodes [27,28,55,76]. Notably, this analysis is typically based on a distance-damped version of the rMI in order to modulate locality in the communication transfer. Differently, information flow (or current flow betweenness) [70,71,74], based on the network Laplacian, allows taking all pathways between source and target into account to highlight nodes carrying effective pathways [21,22,49–51].
- A novel Signal-to-Noise Ratio (SNR) was recently introduced to estimate the efficiency of communication between selected regions of the biomolecules [38,39]. Based on optimal and suboptimal pathways, it quantifies how communication pathways between predefined distant sites (the signal) are favorable over the remaining pathways (the noise) in the network.

Several pieces of software are available to assist in the different stages of network construction and analysis, and they are briefly summarized in Box 1.

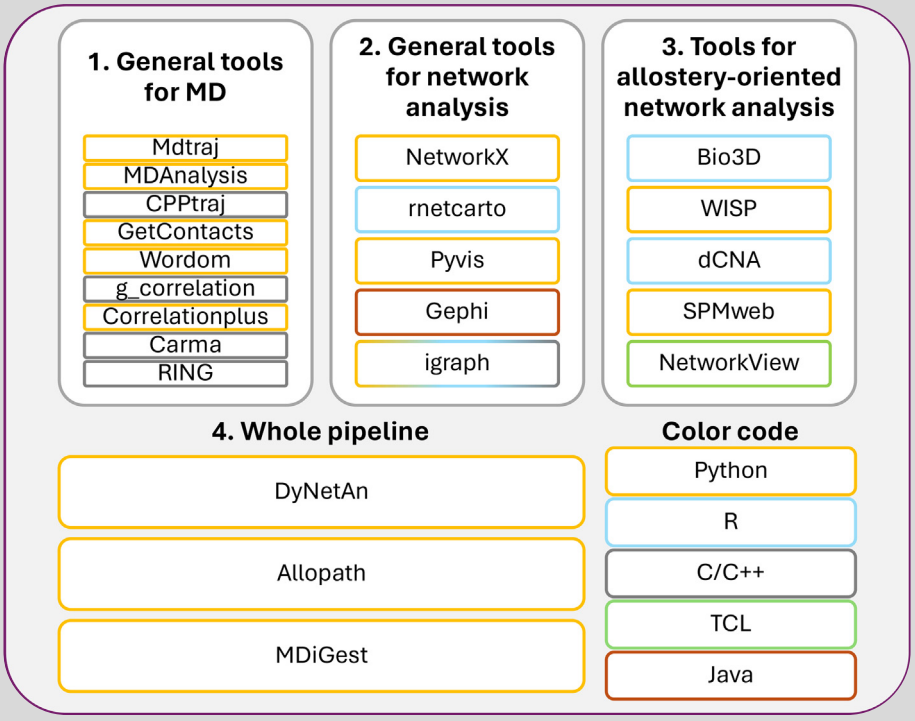
Multi-ensemble networks

Methods based on networks built leveraging multiple conformational ensembles have recently appeared. Here, the communication within the considered biomolecule is first estimated separately through MD simulations performed under different conditions, like the unbound (apo) and ligand-bound (holo) states. The information gathered from the two conformational ensembles is then integrated into a single, multi-ensemble network, with the purpose of capturing alterations associated with the varying condition (or perturbation). Being the allosteric mechanism precisely defined in terms of a transition between ensembles, it has been argued that the analysis of such a multi-ensemble network should be better suited for understanding the functional process, especially for systems undergoing significant conformational changes [77].

Examples are the Dynamical Perturbation Contact Network (DPCN) [78] and difference Contact Network Analysis (dCNA) [79] methods, which have found recent application [26,32–36,54–60]. Both schemes are based on the evaluation of the amount of inter-residue contacts in the separate ensembles. Such information is then conveyed into a network with edges weighted via the difference in contacts between the two ensembles. Suitable thresholds can be then applied to straightforwardly visualize the communication perturbation induced by the changing condition (e.g. effector binding) [78]. Alternatively, standard network analyses, including optimal/suboptimal pathways [36] and node centrality [36], can be used to identify key residues potentially engaged in allosteric processes. An approach for clustering edges while preserving spatial proximity in the network was recently introduced in the context of DPCN, but could be of general use as extendable to any weighted graph [54]. Differently, the consensual framework of dCNA allows a consistent community partition across ensembles, facilitating the interpretation of the allosteric mechanism [79].

We note that both DPCN and dCNA approaches leverage calculation of the number of contacts between residue pairs, rather than correlation via other more commonly used metrics (Table 1). However, the idea is in principle applicable to correlation-based networks too [74], and differences of the whole correlation matrices or node

Box 1. Software used by the references reported in the present review.



1. General-purpose resources to analyze trajectories from MD simulations. This group also includes tools to compute different types of correlations in MD trajectories, not necessarily to study allosteric mechanisms.
2. General-purpose resources devised to construct and analyze networks, not necessarily using MD trajectory data.
3. Tools devised specifically to construct/analyze networks built from MD trajectory data and aimed at investigating allosteric mechanisms.
4. Tools that cover the whole pipeline of investigating allostery from MD simulations using network analysis. Their utilities range from computing correlations to constructing and analyzing the networks.

In the schematic, different colors indicate the programming language on which they are based.

eigenvector centralities were already employed to compare systems under different conditions [20,28,30,55,63].

Conclusion and future outlook

The identification of communication pathways and critical hotspots for signal transduction in biomolecular systems is of paramount importance to unveil the mechanistic signatures of allostery. This is fundamental to allow the rational design of novel therapeutic strategies. The exploitation of the allosteric modulation in drug discovery already gave tangible achievements, with a number of drugs already on the market and others under clinical evaluation [8,80]. This involves historical targets such as kinases, ion channels and GPCRs, and holds great potential for emerging targets such as nuclear receptors and membrane transporters, for which the rich pattern of allosteric mechanisms is gradually being recognized and characterized. Computational approaches can be game-changers here, as the atomistic details of allosteric communication are difficult to assess experimentally. MD simulations allow investigating functional mechanisms with atomistic resolution. In principle, they can be exploited to study the full allosteric transitions. However, using plain (i.e. unbiased) MD, possibly coupled to Markov State Models analysis methods, can result in an overwhelming computational burden, which does not comply with the efficiency requirements of computational drug discovery. On the contrary, enhanced sampling simulations can drive the allosteric transition at a lower computational cost, provided sufficient mechanistic knowledge is available. This can be particularly beneficial when orthogonal degrees of freedom contribute to the complexity of the

process, such as lipid dynamics affecting the functionality of membrane proteins. In this respect, AI-based detection of features relevant for the process can be extraordinarily valuable in overcoming sampling limitations and improving the characterization of the transition. In this review, we focused on methods that allow investigating allosteric communication by studying the distinct conformational dynamics at the endpoints, such as the ligand-bound and unbound states. These approaches combine network analysis with MD simulations, and are gradually becoming of routine use. Our survey to assemble this review highlighted the modular nature of these methods in analyzing allostery, comprising different metrics, distinct choices of parameters for network construction, and diverse possibilities for the analyses of the produced networks. Most remarkably, all these options can be integrated into different combinations. This is reflected by the wealth of data presented in Table 1, which should, however, only be taken as a guide, as it cannot exhaustively capture the granularity of each individual contribution. While we consider this versatile picture highly exciting, our desire would be an effort towards standardization of the procedures. Converging to a common framework for a unified description of the essential steps undertaken in network construction/analysis and details on critical parameters could be extremely valuable to make practices easier to reproduce, evaluate, and, possibly, compare. In this respect, we believe the recent surge of dedicated resources that are openly available, including the DyNetAn [69], Allopath [70], and MDiGest [75] packages, is extremely valuable in fostering reproducibility and accessibility in the wider scientific community. Finally, integrating the emerging AI methods, which allow improving the conformational sampling, the analysis of trajectories, and the network analysis itself, can disclose exciting scenarios toward actionable allosteric drug discovery.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data were used for the research described in the article.

Acknowledgements

Mattia Bernetti acknowledges funding by the project "National Center for Gene Therapy and Drugs based on RNA Technology" (CN00000041), financed by NextGenerationEU PNRR MUR – M4C2 – Action 1.4 – Call "Potenziamento strutture di ricerca e di campioni nazionali di R&S" (CUP: J33C22001130001).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Changeux J-P: **The origins of allostery: from personal memories to material for the future**. *J Mol Biol* 2013, **425**: 1396–1406.

2. Monod J, Wyman J, Changeux J-P: **On the nature of allosteric transitions: a plausible model**. *J Mol Biol* 1965, **12**:88–118.

3. Koshland DE, Némethy G, Filmer D: **Comparison of experimental binding data and theoretical models in proteins containing subunits**. *Biochemistry* 1966, **5**:365–385.

4. Cooper A, Dryden D: **Allostery without conformational change: a plausible model**. *Eur Biophys J* 1984, **11**:103–109.

5. Guo J, Zhou H-X: **Protein allostery and conformational dynamics**. *Chem Rev* 2016, **116**:6503–6515.

6. Gunasekaran K, Ma B, Nussinov R: **Is allostery an intrinsic property of all dynamic proteins?** *Proteins Struct Funct Bioinforma* 2004, **57**:433–443.

7. Gibbs AC: **Elements and modulation of functional dynamics**. *J Med Chem* 2014, **57**:7819–7837.

8. Han B, Salituro FG, Blanco M-J: **Impact of allosteric modulation in drug discovery: innovation in emerging chemical modalities**. *ACS Med Chem Lett* 2020, **11**:1810–1819.

9. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE: **Emerging computational methods for the rational discovery of allosteric drugs**. *Chem Rev* 2016, **116**: 6370–6390.

10. Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL: **Investigating cryptic binding sites by molecular dynamics simulations**. *Acc Chem Res* 2020, **53**:654–661.

11. Colombo G: **Computing allostery: from the understanding of biomolecular regulation and the discovery of cryptic sites to molecular design**. *Curr Opin Struct Biol* 2023, **83**, 102702.

12. Vendruscolo M, Dokholyan NV, Paci E, Karplus M: **Small-world view of the amino acids that play a key role in protein folding**. *Phys Rev E* 2002, **65**, 061910.

13. * Nerín-Fonz F, Cournia Z: **Machine learning approaches in predicting allosteric sites**. *Curr Opin Struct Biol* 2024, **85**, 102774.
This review is an excellent survey of recent applications of machine learning approaches to investigate allostery, focusing in particular on methods exploiting machine learning for allosteric modulator binding site prediction.

14. Agajanian S, Alshahrani M, Bai F, Tao P, Verkhivker GM: **Exploring and learning the universe of protein allostery using artificial intelligence augmented biophysical and computational approaches**. *J Chem Inf Model* 2023, **63**:1413–1428.

15. Sethi A, Eargle J, Black AA, Luthey-Schulten Z: **Dynamical networks in tRNA: protein complexes**. *Proc Natl Acad Sci* 2009, **106**:6620–6625.

16. Naseem-Khan S, Berger MB, Leddin EM, Maghsoud Y, Cisneros GA: **Impact of Remdesivir incorporation along the primer strand on SARS-CoV-2 RNA-dependent RNA polymerase**. *J Chem Inf Model* 2022, **62**:2456–2465.

17. Ginex T, Marco-Marín C, Wieczór M, Mata CP, Krieger J, Ruiz-Rodriguez P, López-Redondo ML, Francés-Gómez C, Melero R, Sánchez-Sorzano CÓ, *et al.*: **The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: the case of the spike A222V mutation**. *PLoS Pathog* 2022, **18**, e1010631.

18. Tajima N, Simorowski N, Yovanno RA, Regan MC, Michalski K, Gómez R, Lau AY, Furukawa H: **Development and characterization of functional antibodies targeting NMDA receptors**. *Nat Commun* 2022, **13**:923.

19. * Krishnan K, Tian H, Tao P, Verkhivker GM: **Probing conformational landscapes and mechanisms of allosteric communication in the functional states of the ABL kinase domain using multiscale simulations and network-based mutational**

profiling of allosteric residue potentials. *J Chem Phys* 2022, **157**.
The authors study the allosteric communication in the ABL kinase domain by constructing a network from both correlation, estimated from MD simulations, and co-evolutionary residue couplings. Interestingly, they introduce a mutational profiling approach to estimate the perturbation of the allosteric communication induced by systematic modification of protein residues. The approach may have interesting application in a drug discovery perspective, e.g. to investigate possible alteration in communication pathways associated with mutations inducing drug resistance.

20. Liu J, Amaral LA, Keten S: **A new approach for extracting information from protein dynamics**. *Proteins Struct Funct Bioinforma* 2023, **91**:183–195.

21. ** Kinnebrew M, Woolley RE, Ansell TB, Byrne EF, Frigui S, Luchetti G, Sircar R, Nachtergaele S, Mydock-McGrane L, Krishnan K, *et al.*: **Patched 1 regulates smoothened by controlling sterol binding to its extracellular cysteine-rich domain**. *Sci Adv* 2022, **8**, eabm5563.
The authors investigate activity modulation in the Smoothened (SMO) protein by cholesterol binding at two distant sites, located respectively in the extracellular cysteine-rich domain and the transmembrane domain. Using crystallography, mutational experiments and network analysis, specifically information flow, they describe the allosteric communication between the two sites and propose a model for SMO activation based on sterol occupancy.

22. ** Zhang G, Xu X, Jia Z, Geng Y, Liang H, Shi J, Marras M, Abella C, Magleby KL, Silva JR, *et al.*: **An allosteric modulator activates BK channels by perturbing coupling between Ca2+ binding and pore opening**. *Nat Commun* 2022, **13**:6784.
The authors integrate in silico studies with electrophysiological and mutational experiments to identify a small molecule that modulates an allosteric mechanism involved in activation of BK channels. The ligand binding mode identified via in silico screening is used to initiate MD simulations, which are then used to perform network analysis, specifically information flow. The study is a remarkable example of coupling modeling approaches with network analysis in a drug discovery context.

23. Kihn K, Marchiori E, Spagnolli G, Boldrini A, Terruzzi L, Lawrence DA, Gershenson A, Faccioli P, Wintrode PL: **Long-range allostery mediates the regulation of plasminogen activator inhibitor-1 by cell adhesion factor vitronectin**. *J Biol Chem* 2022, **298**, 102652.

24. Sun B, Kekenes-Huskey PM: **Calmodulin's interdomain linker is optimized for dynamics signal transmission and calcium binding**. *J Chem Inf Model* 2022, **62**:4210–4221.

25. Calvó-Tusell C, Maria-Solano MA, Osuna S, Feixas F: **Time evolution of the millisecond allosteric activation of imidazole glycerol phosphate synthase**. *J Am Chem Soc* 2022, **144**:7146–7159.

26. Gheeraert A, Vuillon L, Chaloin L, Moncorgé O, Very T, Perez S, Leroux V, Chauvot de Beauchêne I, Mias-Lucquin D, Devignes M-D, *et al.*: **Singular interface dynamics of the SARS-CoV-2 delta variant explained with contact perturbation analysis**. *J Chem Inf Model* 2022, **62**:3107–3122.

27. Kornev AP, Aoto PC, Taylor SS: **Calculation of centralities in protein kinase A**. *Proc Natl Acad Sci* 2022, **119**, e2215420119.

28. Maschietto F, Gheeraert A, Piazzi A, Batista VS, Rivalta I: **Distinct allosteric pathways in imidazole glycerol phosphate synthase from yeast and bacteria**. *Biophys J* 2022, **121**:119–130.

29. Yang T, Han L, Huo S: **Dynamics and allosteric information pathways of unphosphorylated c-Cbl**. *J Chem Inf Model* 2022, **62**:6148–6159.

30. Ray D, Quijano RN, Andricioaei I: **Point mutations in SARS-CoV-2 variants induce long-range dynamical perturbations in neutralizing antibodies**. *Chem Sci* 2022, **13**:7224–7239.

31. Santos NP, Santos LH, Torquato Quezado de Magalhães M, Lei J, Hilgenfeld R, Salgado Ferreira R, Bleicher L: **Characterization of an allosteric pocket in zika virus NS2B-NS3 protease**. *J Chem Inf Model* 2022, **62**:945–957.

32. Barbera N, Granados ST, Vanoye CG, Abramova TV, Kulbak D, Ahn SJ, George AL, Akpa BS, Levitan I: **Cholesterol-induced suppression of Kir2 channels is mediated by decoupling at the inter-subunit interfaces**. *iScience* 2022, **25**.

33. Fung Y-H, Kong W-P, Leung ASL, Du R, So P-K, Wong W-L, Leung Y-C, Chen YW, Wong K-Y: **NDM-1 Zn1-binding residue His116 plays critical roles in antibiotic hydrolysis**. *Biochim Biophys Acta BBA Proteins Proteom* 2022, **1870**, 140833.

34. Li M, Wang Y, Fan J, Zhuang H, Liu Y, Ji D, Lu S: **Mechanistic insights into the long-range allosteric regulation of KRAS via neurofibromatosis type 1 (NF1) scaffold upon SPRED1 loading**. *J Mol Biol* 2022, **434**, 167730.

35. Kumutima J, Yao X-Q, Hamelberg D: **Post-translational modifications of cyclophilin D fine-tune its conformational dynamics and activity: implications for its mitochondrial function**. *J Phys Chem B* 2022, **126**:10844–10853.

36. Yao X-Q, Hamelberg D: **Residue–residue contact changes during functional processes define allosteric communication pathways**. *J Chem Theory Comput* 2022, **18**:1173–1187.

37. Costa F, Guardiani C, Giacomello A: **Molecular dynamics simulations suggest possible activation and deactivation pathways in the hERG channel**. *Commun Biol* 2022, **5**:165.

38. * Sinha S, Molina Vargas AM, Arantes PR, Patel A, O'Connell MR, Palermo G: **Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a**. *Nucleic Acids Res* 2023, **52**:906–920.
The authors use extensive MD simulations, in the hundred microseconds time scale, and network analysis to study the allosteric activation mechanism in the emerging CRISPR-Cas13a system. Leveraging a novel Signal-to-Noise Ratio estimate of the information transfer in the network, they identify key residues for allosteric communication. To validate predictions, they experimentally study Cas13a variants using fluorescent RNA trans-cleavage assays.

39. Molina Vargas AM, Sinha S, Osborn R, Arantes PR, Patel A, Dewhurst S, Hardy DJ, Cameron A, Palermo G, O'Connell MR: **New design strategies for ultra-specific CRISPR-Cas13a-based RNA detection with single-nucleotide mismatch sensitivity**. *Nucleic Acids Res* 2023, **52**:921–939.

40. Liu D, Zhang H, Huang Y-P, Gao YQ: **Investigating the activation mechanism differences between human and mouse cGAS by molecular dynamics simulations**. *J Phys Chem B* 2023, **127**:5034–5045.

41. Yovanno RA, Yu A, Wied TJ, Lau AY: **Phosphorylation at Ser65 modulates ubiquitin conformational dynamics**. *Structure* 2023, **31**:884–890.e2.

42. * Aguti R, Bernetti M, Bosio S, Decherchi S, Cavalli A: **On the allosteric puzzle and pocket crosstalk through computational means**. *J Chem Phys* 2023, **158**, 165101.
The authors compare methods falling within the category cured in the present review with other MD-based computational approaches used in the context of allosteric communication, namely pocket cross-talk and distance fluctuations. Remarkably, the latter approach is framed into a network representation in order to enable a more straightforward comparison with the other methods. In all cases, the communication is evaluated between pockets rather than individual residues.

43. Li H, Urs NM, Horenstein N: **Computational insights into ligand–induced G protein and β-arrestin signaling of the dopamine D1 receptor**. *J Comput Aided Mol Des* 2023, **37**:227–244.

44. Spinello A, Janos P, Rozza R, Magistrato A: **Cancer-related mutations alter RNA-driven functional cross-talk underlying premature-messenger RNA recognition by splicing factor SF3B**. *J Phys Chem Lett* 2023, **14**:6263–6269.

45. Dube N, Khan SH, Sasse R, Okafor CD: **Identification of an evolutionarily conserved allosteric network in steroid receptors**. *J Chem Inf Model* 2023, **63**:571–582.

46. Konovalov KA, Wu C-G, Qiu Y, Balakrishnan VK, Parihar PS, O'Connor MS, Xing Y, Huang X: **Disease mutations and phosphorylation alter the allosteric pathways involved in autoinhibition of protein phosphatase 2A**. *J Chem Phys* 2023, **158**.

47. Soya N, Xu H, Roldan A, Yang Z, Ye H, Jiang F, Premchandar A, Veit G, Cole SP, Kappes J: **Folding correctors can restore CFTR posttranslational folding landscape by allosteric domain–domain coupling**. *Nat Commun* 2023, **14**:6868.

48. Dayananda A, Dennison T, Fonseka HYY, Avestan MS, Wang Q, Tehver R, Stan G: **Allosteric communication in the gating mechanism for controlled protein degradation by the bacterial ClpP peptidase**. *J Chem Phys* 2023, **158**, 125101.

49. Janaszkiewicz A, Tóth Á, Faucher Q, Arnion H, Védrenne N, Barin-Le Guellec C, Marquet P, Di Meo F: **Substrate binding and lipid-mediated allostery in the human organic anion transporter 1 at the atomic-scale**. *Biomed Pharmacother* 2023, **160**, 114342.

50. Tóth Á, Janaszkiewicz A, Crespi V, Di Meo F: **On the interplay between lipids and asymmetric dynamics of an NBS degenerate ABC transporter**. *Commun Biol* 2023, **6**:149.

51. Tóth Á, Crespi V, Janaszkiewicz A, Di Meo F: **Computational and structural insights into the pre-and post-hydrolysis states of bovine multidrug resistance-associated protein 1**. *Basic Clin Pharmacol Toxicol* 2023, **133**:508–525.

52. Cheng RL, Quirante JC, Vargas LEZ, Gatchalian AF, Nellas RB: **Complementary pocket and network-based approach to search for spike protein allosteric pocket sites**. *ACS Omega* 2023, **8**:45313–45325.

53. Omotuyi OI, Nash O, Metibemu SD, Iwegbulam CG, Olatunji OM, Agbebi E, Falade OC: **Dihydroartemisinin binds human PI3K-affinity pocket and forces flat conformation in P-loop MET783: a molecular dynamics study**. *Comput Toxicol* 2023, **27**, 100281.

54. Gheeraert A, Lesieur C, Batista VS, Vuillon L, Rivalta I: **Connected component analysis of dynamical perturbation contact networks**. *J Phys Chem B* 2023, **127**:7571–7580.

55. * Maschietto F, Morzan UN, Tofoleanu F, Gheeraert A, Chaudhuri A, Kyro GW, Nekrasov P, Brooks B, Loria JP, Rivalta I, *et al.*: **Turning up the heat mimics allosteric signaling in imidazole-glycerol phosphate synthase**. *Nat Commun* 2023, **14**:2239.
In this study, MD simulations and network analysis are exploited together with NMR experiments to study the effect of increases in temperature on allosteric communication in imidazole glycerol phosphate synthase. The study, by combining computational and experimental approaches, provides an atomistic picture of temperature-dependent allostery.

56. Chen YW, Kong W-P, Wong K-Y: **The structural integrity of the membrane-embedded bacterial division complex FtsQBL studied with molecular dynamics simulations**. *Comput Struct Biotechnol J* 2023, **21**:2602–2612.

57. Ouedraogo D, Souffrant M, Yao X-Q, Hamelberg D, Gadda G: **Non-active site residue in loop L4 alters substrate capture and product release in d-arginine dehydrogenase**. *Biochemistry* 2023, **62**:1070–1081.

58. Pegram L, Riccardi D, Ahn N: **Activation loop plasticity and active site coupling in the MAP kinase, ERK2**. *J Mol Biol* 2023, **435**, 168309.

59. Souffrant M, Yao X-Q, Hamelberg D: **Evolving mutational buildup in HIV-1 protease shifts conformational dynamics to gain drug resistance**. *J Chem Inf Model* 2023, **63**:3892–3902.

60. Yu J, Yan C, Dodd T, Tsai C-L, Tainer JA, Tsutakawa SE, Ivanov I: **Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases**. *Nat Commun* 2023, **14**:2758.

61. Kelly MS, Macke AC, Kahawatte S, Stump JE, Miller AR, Dima RI: **The quaternary question: determining allostery in spastin through dynamics classification learning and bioinformatics**. *J Chem Phys* 2023, **158**, 125102.

62. Crean RM, Slusky JS, Kasson PM, Kamerlin SCL: **KIF – key interactions finder: a program to identify the key molecular interactions that regulate protein conformational changes**. *J Chem Phys* 2023, **158**, 144114.

63. Srivastava A, Idriss H, Homouz D: **Structural insights into phosphorylation-mediated polymerase function loss for DNA polymerase β bound to gapped DNA**. *Int J Mol Sci* 2023, **24**:8988.

64. Jani V, Sonavane U, Sawant S: **Structural insights into the activation mechanism of phosphoinositide 3-kinase alpha**. *Comput Biol Chem* 2024, **108**, 107994.

65. ** Bassetto Jr CA, Costa F, Guardiani C, Bezanilla F, Giacomello A: **Noncanonical electromechanical coupling paths in cardiac hERG potassium channel**. *Nat Commun* 2023, **14**:1110.
The authors study the allosteric communication relevant for the gating mechanism of the hERG potassium channel. By using network analysis, they identify key residues for the process, suggesting a non-canonical electromechanical pathway. The relevance of the predicted residues is confirmed through mutagenesis and electrophysiological experiments, providing novel insight into activation and inactivation mechanisms of the hERG channel.

66. Costa F, Ocello R, Guardiani C, Giacomello A, Masetti M: **Integrated approach including docking, MD simulations, and network analysis highlights the action mechanism of the cardiac hERG activator RPR260243**. *J Chem Inf Model* 2023, **63**:4888–4899.

67. Rivalta I, Sultan MM, Lee N-S, Manley GA, Loria JP, Batista VS: **Allosteric pathways in imidazole glycerol phosphate synthase**. *Proc Natl Acad Sci* 2012, **109**:E1428–E1436.

68. Lange OF, Grubmüller H: **Generalized correlation for biomolecular dynamics**. *Proteins Struct Funct Bioinforma* 2006, **62**:1053–1061.

69. Melo MC, Bernardi RC, De La Fuente-Nunez C, Luthey-Schulten Z: **Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories**. *J Chem Phys* 2020, **153**, 134104.

70. Westerlund AM, Fleetwood O, Pérez-Conesa S, Delemotte L: **Network analysis reveals how lipids and other cofactors influence membrane protein allostery**. *J Chem Phys* 2020, **153**, 141103.

71. Kang PW, Westerlund AM, Shi J, White KM, Dou AK, Cui AH, Silva JR, Delemotte L, Cui J: **Calmodulin acts as a state-dependent switch to control a cardiac potassium channel opening**. *Sci Adv* 2020, **6**, eabd6798.

72. Van Wart AT, Durrant J, Votapka L, Amaro RE: **Weighted implementation of suboptimal paths (WISP): an optimized algorithm and tool for dynamical network analysis**. *J Chem Theory Comput* 2014, **10**:511–517.

73. Dodd T, Yao X-Q, Hamelberg D, Ivanov I: **Subsets of adjacent nodes (SOAN): a fast method for computing suboptimal paths in protein dynamic networks**. *Mol Phys* 2021, **119**, e1893847.

74. Botello-Smith WM, Luo Y: **Robust determination of protein allosteric signaling pathways**. *J Chem Theory Comput* 2019, **15**:2116–2126.

75. ** Maschietto F, Allen B, Kyro GW, Batista VS: **MDiGest: a Python package for describing allostery from molecular dynamics simulations**. *J Chem Phys* 2023, **158**.
The authors present a python package that allows studying allosteric mechanisms from MD simulations through network analysis. This very recent package, named MDiGest, covers the whole pipeline of the procedure, from computing correlations to constructing and analyzing the networks. Interestingly, it implements a variety of non-standard features/metrics, such as torsional angles and electrostatic energies.

76. Negre CF, Morzan UN, Hendrickson HP, Pal R, Lisi GP, Loria JP, Rivalta I, Ho J, Batista VS: **Eigenvector centrality for characterization of protein allosteric pathways**. *Proc Natl Acad Sci* 2018, **115**:E12201–E12208.

77. ** Yao X-Q, Hamelberg D: **From distinct to differential conformational dynamics to map allosteric communication pathways in proteins**. *J Phys Chem B* 2022, **126**:2612–2620.
In this perspective, the authors provide a thorough overview on the basic concepts of allostery. We recommend it both as an introductory read to newcomers in the field and to dive into more recent and advanced developments, such as the methodology we referred to as multi-ensemble network in the present review.

78. Gheeraert A, Pacini L, Batista VS, Vuillon L, Lesieur C, Rivalta I: **Exploring allosteric pathways of a v-type enzyme with dynamical perturbation networks**. *J Phys Chem B* 2019, **123**:3452–3461.

79. Yao X-Q, Momin M, Hamelberg D: **Elucidating allosteric communications in proteins with difference contact network analysis**. *J Chem Inf Model* 2018, **58**:1325–1330.

80. Sheik Amamuddy O, Veldman W, Manyumwa C, Khairallah A, Agajanian S, Oluyemi O, Verkhivker GM, Tastan Bishop Ö: **Integrated computational approaches and tools for allosteric drug discovery**. *Int J Mol Sci* 2020, **21**:847.