Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Protein contact network topology: a natural language for allostery

Luisa Di Paola[1] and Alessandro Giuliani[2]

Protein molecules work as a whole, so that any local perturbation may resonate on the entire structure: allostery deals with this general property of protein molecules. It is worth noting a perturbation does not necessarily involve a conformational change but, more generally, it travels across the structure as an 'energy signal'. The atomic interactions within the network provide the structural support for this 'signaling highway'. Network descriptors, capturing network signaling efficiency, explain allostery in terms of signal transmission.In this review we will survey the key applications of graph theory to protein allostery. The complex network approach introduces a new perspective in biochemistry; as for applications, the development of new drugs relying on allosteric effects (allo-network drugs) represents a promising avenue of contact network formalism.

Addresses
[1] Faculty of Engineering, Università CAMPUS Bio-Medico, Via A. del Portillo, 21, 00128 Rome, Italy
[2] Environment and Health Department, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Roma, Italy

Corresponding author: Di Paola, Luisa (l.dipaola@unicampus.it)

Current Opinion in Structural Biology 2015, **31**:43–48

This review comes from a themed issue on **Theory and simulation**

Edited by **Claire Lesieur** and **Klaus Schulten**

For a complete overview see the Issue and the Editorial

Available online 18th March 2015

http://dx.doi.org/10.1016/j.sbi.2015.03.001

Introduction

Allostery is a neologism modeled upon Greek language, which has to do with the ability of proteins to transmit a signal from one site to another in response to environmental stimuli. This ability is related to the transmission of information across the protein molecule from a sensor (allosteric) site to the effector (binding) site [1•]. The molecule, hence, perceives ligand binding at distance from the active site, or any other microenvironmental perturbation, like pH changes.

The information transfer across protein molecules was studied along different lines: we will show in the following that complex network analysis allows for a reunification of different mechanisms [2••], providing a promising basis for an innovative pharmacological approach [3–5].

Fifty years ago, Monod, Wyman, and Changeux presented a simple model of allostery (MWC) based on the interaction between distinct sites mediated by protein conformational changes [6]. According to MWC, two (or more) interchangeable conformational states of the protein co-exist in a thermal equilibrium; the states — often termed tense (T) and relaxed (R) — differ in affinity for the ligand molecule and derive from concerted motions of subunits. The ensemble distribution of these states depends upon the binding of small ligand molecules, stabilizing the higher-affinity state. Daniel Koshland and colleagues (KNF — Koshland–Nemethy–Filmer — model) proposed a slightly different view of the process, setting a sequential induced fit paradigm [7]. Both MWC and KNF models require a clear distinction between different conformational states that switch upon binding. In addition to these two models there is also a MWC without conformational changes set forth by Cooper and Dryden [8].

Thermodynamic considerations have offered a framework to the different models of allostery but they did not provide the mechanism of allostery, that is, how do the changes propagate from one place to another. This is exactly where the network approach comes in: 'Complex networks of interacting residues and microdomains in the structures of biomolecular systems underlie the reliable propagation of information from an input signal, such as the concentration of a ligand, to sites that generate the appropriate output signal, such as enzymatic activity' [9].

Binding free energy comprises an enthalpic and an entropic contribution [10]: the fast, local rearrangements around the stable position of single residues correspond to the entropic term, while enthalpy accounts for global and relatively slow motions provoking conformational changes. According to this view, only processes comprising non-negligible enthalpic contributions result into global conformational changes, whereas purely entropic processes occur with no appreciable conformational changes.

The review will show how network approaches to allostery enable to first, distinguish cases with and without conformational changes, second, identify allostery residues and third, find the binding signal transmission routes (communication channels). This noteworthing contact network formalism is based on a strong reduction of structural information, while keeping its essence, so

the functional outcomes of structural data can be revealed in terms of network descriptors.

The topology of protein contact networks (PCNs) and its link with allostery

PCNs catch the essential of signal (and energy) transmission in terms of wiring architecture of the system [11,12•, 13•,14,15] (see Box 1 for details on network descriptors).

In PCNs the shortest paths mediate concerted motions and energy transmission upon ligand binding [16,17]. The topological metrics of shortest paths (minimum number of links separating two residues) is thus the 'actual' metrics for signaling.

In the case of quaternary structures, modules (domains) naturally correspond to single chains and their mutual interactions in allostery have been thoroughly analyzed in eminent case studies [18].

Spectral clustering algorithms allow for protein decomposition into modules [19] (see Box 1). These methods apply to the adjacency matrix, so the partition roots on the interactions between residues (links) and only indirectly on their mutal distances (topological clustering).

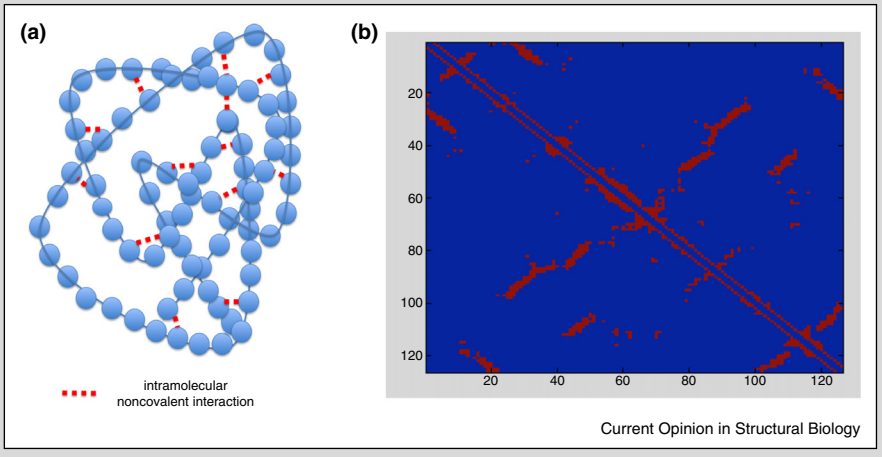
Figure 1 reports the modules (domains) identification in hemoglobin by spectral clustering (panel a), which identifies the four chains (consecutive blocks of different color, 141 residues in chains A and C, 146 in chains B and D). In addition, spectral decomposition highlights inter-module pathways, made of the (few) residues pertaining to a given chain but topologically 'more similar' to other domains ('whiskers' in panel a). Geometrical clustering (k-means), based on Euclidean coordinates of residues, results into clusters almost exactly matching with single chains, missing the functionally active residues responsible for 'intermodule communication' (Figure 1, panel b) [20]. This is a vivid demonstration of the 'added-value' of topological versus geometrical approach to protein structure elucidation.

Guimerà–Amaral cartography [21] allows to define the role of individual residues in terms of 'inter-module' and

Box 1 Protein contact networks.

PCNs describe the intramolecular interaction networks in protein molecules; nodes are the single residues (identified by the corresponding α-carbon) and edges between two nodes exist if their Euclidean distance falls within 4–8 Å range, so to include noncovalent interactions — sensitive to environmental stimuli (see Figure B1a).

Figure B1



Panel **(a)**: the intramolecular non-covalent interactions (red dotted lines) connect spatially close residues and panel **(b)**: the intramolecular interactions network translates into the adjacency matrix, a binary matrix whose elements are non-null (red dots) if the corresponding residues are in contact (both axes of the matricial representation correspond to sequence.

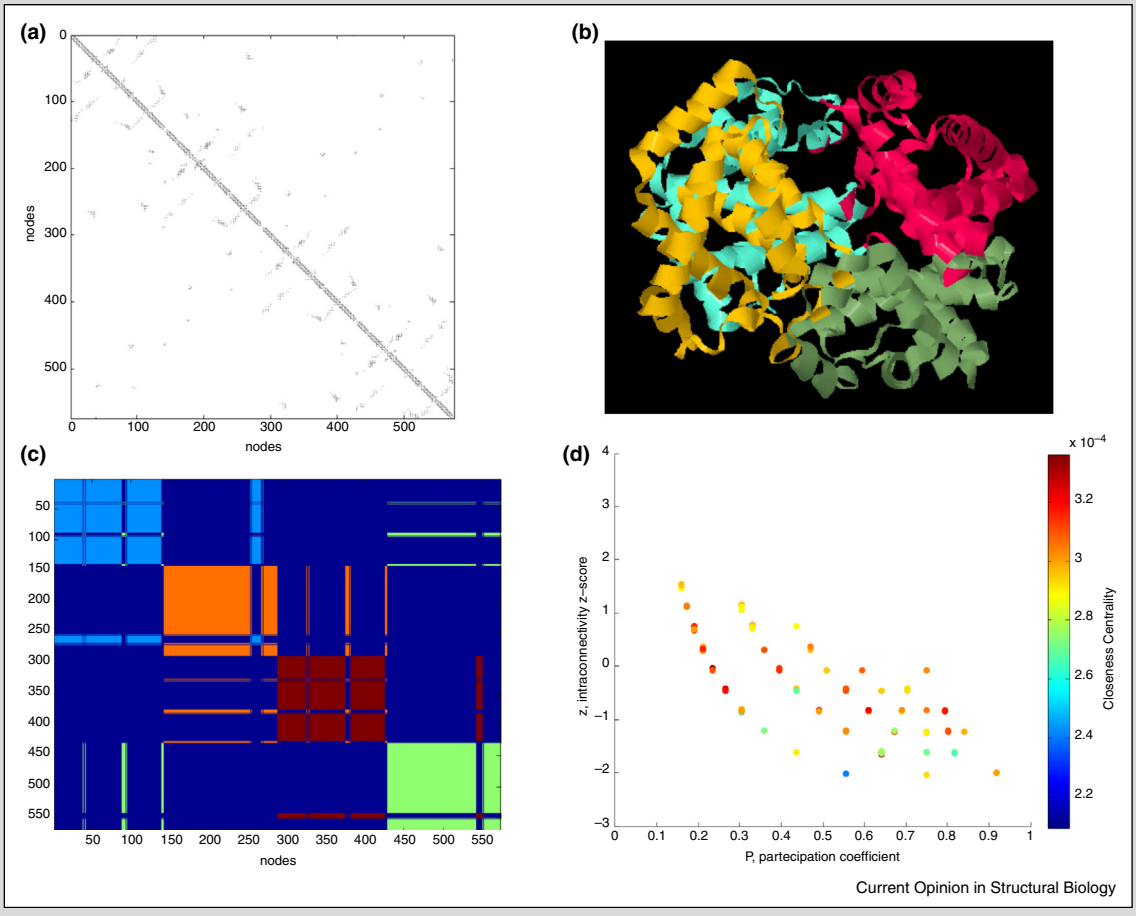
The contact network is mathematically formalized by the adjacency matrix A (Figure B1b), whose generic element A_{ij} is 1 if the ith and jth residues are in contact, 0 otherwise.

The adjacency matrix allows to compute the shortest path between two nodes in the network, which represents the minimum number of links connecting them. The average shortest path is the average length connecting any pair of nodes (residues) in the network.

The node centrality strictly depends on the shortest paths: the closeness centrality is the inverse of fairness, which, for a generic ith node, is the sum of its shortest paths. High centrality residues connect different domains (modules).

Once defined the adjacency matrix, modules in the network are identified by means of the spectral clustering methodology, able to part the network into clusters. The spectral clustering technique operates the space decomposition through the adjacency matrix eigenvalues, so that the partition relies on the topological role of residues in the interaction network, rather than on their spatial positioning (see Figure B2).

Figure B2



Hemoglobin spectral clustering partition. Panel **(a)**: the hemoglobin contact network is represented by the adjacency matrix (dots correspond to contacts between residues); panel **(b)**: modules identified by spectral clustering are in different colors on the ribbon structure representation; panel **(c)**: the clustering color map reports the cluster partition along the sequence; panel **(d)**: the *P*–*z* diagram sketches connectivity in terms of the Guimerà–Amaral.

The nodes (residues) role is specifically addressed by means of two features, describing the attitude of nodes to participate into links within their own community or outside: the *z* (*intraconnectivity z-score*) represents the tendency of nodes to establish contacts within their own community (zero represents the whole community average, positive values an intramodular connectivity higher than the average), whereas *P* (*partecipation coefficient*) describes the proneness of nodes to link with other nodes belonging to different communities (clusters). The Guimerà–Amaral cartography allows to part the *P*–*z* space in regions identifying nodes with different topological roles. The explicit formulas for *P* and *z* are:

$$P_i = 1 - \left(\frac{k_{si}}{k_i}\right)^2 \quad Z_i = \frac{k_i - \bar{k_{si}}}{\sigma_{si}}$$

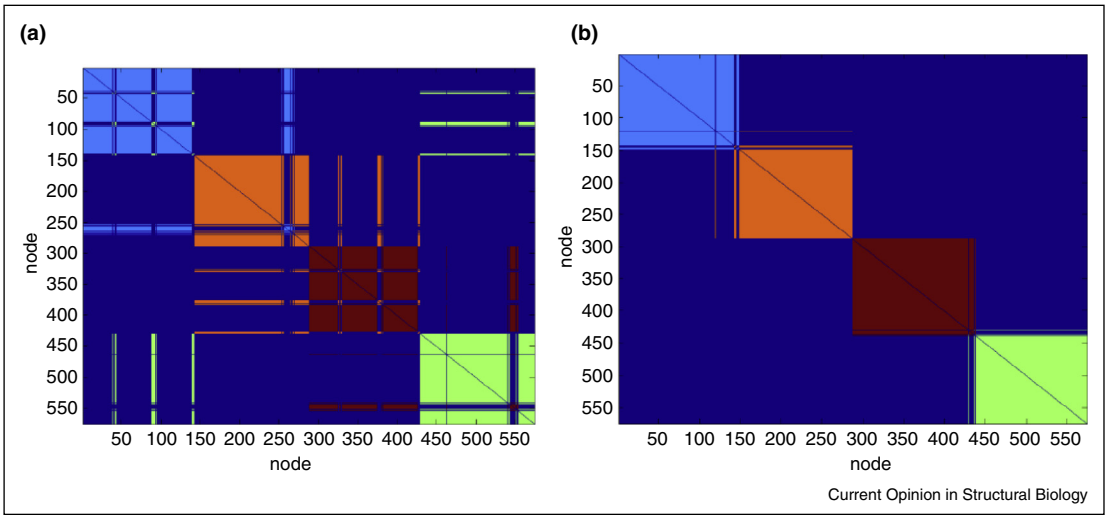
being k_i the *i*th node degree, that is, the number of links *i*th residue is involved into; k_{si} is the intramodule degree of the *i*th node; $\bar{k_{si}}$ and σ_{si} are, respectively, the average and the standard deviation values of the intramodule degree for the module it pertains to.

'intra-module' communication, using two descriptors *P* — participation coefficient, residues with higher values of *P* play a role in inter-module communication and z — general connectivity, nodes with high z and low *P* are mainly endowed with 'intra-module' character, see the Box 1 for details. The relevance of Guimerà–Amaral cartography to allostery is reported in [22••]: allosteric effects are registered in changes of the parameter *P* upon binding (Figure 2).

In allosteric proteins, residues close (in sequence) and in the active site experience *P* shift from non-null to null values upon binding, while in proteins with a low allostery character this change is not so evident.

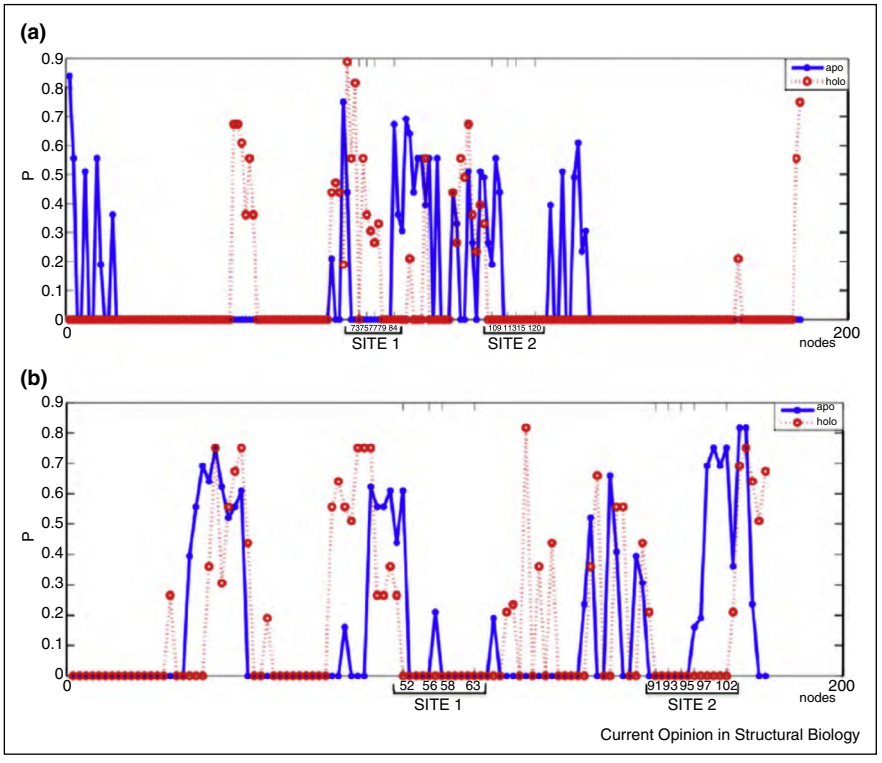
In [23], we reported the superposition of the *P*–z profiles of around 2000 proteins, reinforcing the notion that all the proteins have allosteric properties up to a certain degree [24–26,27•] only varying in the 'amount' of allosteric

Figure 1



Clustering color map of hemoglobin (PDB code 2HHB, 141 residues in chains A and C, 146 in chains B and D): **(a)** spectral clustering (topological metrics); **(b)** *k*-means (Euclidean metrics). Spectral clustering (a) is characterized by horizontal (vertical) lines starting from the cluster core and going forward; they represent the intermodules (chains) links. Geometrical representation (*k*-means), while globally similar, does not allow to catch residues involved in quaternary interactions, crucial for allostery [20].

Figure 2



P values for active site residues in calcium-binding proteins: **(a)** recoverin, an allosteric calcium sensor; **(b)** parvalbumin, a non-allosteric calcium buffer: the value of *P* for recoverin active sites residues (panel a) considerably changes upon binding, while it remains quite steady upon binding in parvalbumin (panel b) [22••].

character. In [22••], the communication pathways across PCNs is faced with an explicit reference to protein modularity. In [28] the communication patterns correlate with the structural and dynamical features of the proteins under analysis: allosteric communication in the different forms of receptor tyrosine kinase, native and mutants, was unraveled by a modular network representation.

Energy flow in proteins [29] provides a dynamic frame to uncouple local and global communication pathways, defining a 'slow' aspecific and a 'fast' directed flow. Considering proteins as structural networks (see Box 1 for details) allows for a natural translation of energy flows in structural terms.

Flow along the fast lane (global motions) is possible since structural networks are arranged as a 'percolation clusters' [29–31]: a specific configuration characterized by network 'channels' (shortest paths) connecting a large number of sites and allowing for short-cuts between otherwise distant protein regions; allostery relies on the 'fast lane' of communication, travelling along such short-cuts.

Network formalism gives an immediate structural counterpart of the energy diffusion inside protein molecules. Inside a protein, energy readily flows between distant regions connected by inter-module contacts (fast lane) and only slowly within modules (slow lane) [29].

In other words, PCNs present both a strong modularity, causing a low thermal dissipation (heat is kept into modules by the richness of dead ends pathways slowing down the spreading of energy [29]) and inter-modules shortcuts, allowing for a rapid and efficient communication between distant sites responsible for allosteric effect.

Network view allows to reconcile the MWC and KND allostery models into a more general 'ensemble selection' frame [32] residues most crucial for allosteric signal transmission show the highest closeness centrality in the interaction network (see the Box 1). del Sol *et al.* [33] gave an experimental proof of this interpretation by means of mutational studies involving central residues and influencing the R–T configuration balance.

PCNs are 'percolation clusters', that is, they have the minimum number of edges guaranteeing the complete connectivity of the network, so allowing the signal spreading across the networks with minimum energy cost [5]. There are 'Multiple pre-existing pathways' [34] allowing for allosteric signaling, corresponding to different conformations in the allostery ensembles [32]; perturbation events simply shift the pre-existing ensemble distribution [27•].

Conclusions

The topology of PCNs promises to play a key role in linking protein structure and function. Allosteric effects are a crucial vantage point to look at protein structure–function relation, because they are the way in which proteins react to their microenvironment. The topological approach to allostery provides a quantitative framework for a new perspective to pharmacology. The seminal work of Csermely *et al.* [5] explicits the innovative idea of 'allo-network drugs', designed under the network formalism guidance [35]. Moreover, the role of allosteric regulation in disease comes out of the allo-network properties of proteins, whose regulatory character faces different scales (from single molecule to the whole cell [36•]).

The biophysical framework to consolidate this approach comes from the development of a suitable methodology to identify the energy flows in protein structures [29], so to forecast the allosteric pathways upon binding [37]. This will allow to conjugate the energy landscape perspective [38] with the biological activity, identifying evolutionarily conserved network of residues responsible of the allosteric response [39], and, at least in perspective, to design drug candidates able to initiate such allosteric cascade [3].

Conflict of interest

Nothing declared.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. • Hilser VJ, Wrabl JO, Motlagh HN: **Structural and energetic basis of allostery**. *Annu Rev Biophys* 2012, **41**:585-609.
This work relies on the ensemble distribution of conformational states linking the allosteric properties to the conformational free energies of binding.

2. •• Tsai C-J, Nussinov R: **A unified view of how allostery works**. *PLoS Comput Biol* 2014, **10**:e1003394.
This work reports a wide evidence of how the thermodynamics of conformational ensembles re-unificates the different molecular theories of allostery, overcoming the hystorical landmark of allostery necessarily linked to wide conformational changes.

3. Szilágyi A, Nussinov R: **Allo-network drugs: extension of the allosteric drug concept to protein–protein interaction and signaling networks**. *Curr Top Med Chem* 2013, **13**:64-77.

4. Lu S, Li S, Zhang J: **Harnessing allostery a novel approach to drug discovery**. *Med Res Rev* 2014, **34**:1242-1285.

5. Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R: **Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review**. *Pharmacol Ther* 2013, **138**:333-408.
A very complete review paper extending the allostery concept, and in general the graph-theory based approach to biological regulation at large. This paper is suggested as an introduction to scientists wishing to apply complex network approach in biology.

6. Changeux J-P: **Allostery and the Monod–Wyman–Changeux model after 50 years**. *Annu Rev Biophys* 2012, **41**:103-133.

7. Changeux JP: **50 years of allosteric interactions: the twists and turns of the models**. *Nat Chem Biol* 2013, **14**:819-829.

8. Cooper A, Dryden DT: **Allostery without conformational change. A plausible model**. *Eur Biophys J* 1984, **11**:103-109.

9. LeVine MV, Weinstein H: **NbIT — a new information theory-based analysis of allosteric mechanisms reveals residues that underlie function in the leucine transporter LeuT**. *PLoS Comput Biol* 2014, **10**:e1003603-e1003615.

10. Homans SW: **Probing the binding entropy of ligand–protein interactions by NMR**. *ChemBioChem* 2005, **6**:1585-1591.

11. Mitternacht S, Berezovsky IN: **A geometry-based generic predictor for catalytic and allosteric sites**. *Protein Eng Des Sel* 2011, **24**:405-409.

12. • Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A: **Protein contact networks: an emerging paradigm in chemistry**. *Chem Rev* 2013, **113**:1598-1613.
A primer on PCNs and graph theoretical methods.

13. • Blacklock K, Verkhivker GM: **Allosteric regulation of the Hsp90 dynamics and stability by client recruiter cochaperones: protein structure network modeling**. *PLoS ONE* 2014, **9**:e86547.
This work reports an interesting case study of the application of the PCNs formalism to elucidate the allosteric mechanism and single out functional residues.

14. Yan W, Zhou J, Sun M, Chen J, Hu G, Shen B: **The construction of an amino acid network for understanding protein structure and function**. *Amino Acids* 2014, **46**:1419-1439.

15. Hu G, Zhou J, Yan W, Chen J, Shen B: **The topology and dynamics of protein complexes: insights from intra-molecular network theory**. *Curr Protein Pept Sci* 2013, **14**:121-132.

16. Bastolla U, Porto M, Roman HE: **The emerging dynamic view of proteins: protein plasticity in allostery, evolution and self-assembly**. *Biochim Biophys Acta* 2013, **1834**:817-819.

17. Berezovsky IN: **Thermodynamics of allostery paves a way to allosteric drugs**. *Biochim Biophys Acta* 2013, **1834**:830-835.

18. Bahar I, Chennubhotla C, Tobi D: **Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation**. *Curr Opin Struct Biol* 2007, **17**:633-640.

19. Kannan N, Vishveshwara S: **Identification of side-chain clusters in protein structures by a graph spectral method**. *J Mol Biol* 1999, **292**:441-464.

20. Tasdighian S, Di Paola L, De Ruvo M, Paci P, Santoni D, Palumbo P, Mei G, Di Venere A, Giuliani A: **Modules identification in protein structures: the topological and geometrical solutions**. *J Chem Inf Model* 2014, **54**:159-168.

21. Guimerà R, Amaral L: **Functional cartography of complex metabolic networks**. *Nature* 2005, **433**:895-900.

22. •• De Ruvo M, Giuliani A, Paci P, Santoni D: **Shedding light on protein–ligand binding by graph theory: the topological nature of allostery**. *Biophys Chem* 2012, **165–166**:21-29.
The paper deals with the allosteric response in terms of topological descriptors of the corresponding PCNs. The modularity of the protein networks is the key to distinguish between allosteric and non-allosteric proteins.

23. Krishnan A, Zbilut JP, Tomita M, Giuliani A: **Proteins as networks: usefulness of graph theory in protein science**. *Curr Protein Pept Sci* 2008, **9**:28-38.

24. Ma B, Tsai C-J, Haliloğlu T, Nussinov R: **Dynamic allostery linkers are not merely flexible**. *Structure* 2011, **19**:907-917.

25. Gunasekaran K, Ma B, Nussinov R: **Is allostery an intrinsic property of all dynamic proteins?** *Proteins* 2004, **57**:433-443.

26. Tsai C-J, del Sol A, Nussinov R: **Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms**. *Mol Biosyst* 2009, **5**:207-216.

27. • Nussinov R, Ma B, Tsai C-J: **Multiple conformational selection and induced fit events take place in allosteric propagation**. *Biophys Chem* 2014, **186**:22-30.
This review summarizes the interpretation of allosteric regulation in terms of ensemble population shift.

28. Laine E, Auclair C, Tchertanov L: **Allosteric communication across the native and mutated KIT receptor tyrosine kinase**. *PLoS Comput Biol* 2012, **8**:e1002661.

29. Leitner DM: **Energy flow in proteins**. *Annu Rev Phys Chem* 2008, **59**:233-259.

30. Morita H, Takano M: **Residue network in protein native structure belongs to the universality class of a three-dimensional critical percolation cluster**. *Phys Rev E* 2009, **79**:020901.

31. Deb D, Vishveshwara S, Vishveshwara S: **Understanding protein structure from a percolation perspective**. *Biophys J* 2009, **97**:1787-1794.

32. Motlagh HN, Wrabl JO, Li J, Hilser VJ: **The ensemble nature of allostery**. *Nature* 2014, **508**:331-339.

33. del Sol A, Fujihashi H, Amoros D, Nussinov R: **Residues crucial for maintaining short paths in network communication mediate signaling in proteins**. *Mol Syst Biol* 2006, **2**.

34. del Sol A, Tsai C-J, Ma B, Nussinov R: **The origin of allosteric functional modulation: multiple pre-existing pathways**. *Structure* 2009, **17**:1042-1050.

35. Nussinov R, Tsai C-J: **The design of covalent allosteric drugs**. *Annu Rev Pharmacol Toxicol* 2014, **55**:249-267.

36. • Nussinov R, Tsai C-J: **Allostery in disease and in drug discovery**. *Cell* 2013, **153**:293-305.
This work surveys the main applications of the allosteric mechanism studies in the clinical practice and pharmacological sciences. Understanding the molecular triggers of the allosteric response represents the starting step for new drugs (orthosteric drugs).

37. Szalay KZ, Csermely P: **Perturbation centrality and turbine a novel centrality measure obtained using a versatile network dynamics tool**. *PLoS ONE* 2013, **8**:e78059-e78111.

38. Swain J, Gierash L: **The changing landscape of protein allostery**. *Curr Opin Struct Biol* 2006, **16**:102-108.

39. Süel GM, Lockless SW, Wall MA, Ranganathan R: **Evolutionarily conserved networks of residues mediate allosteric communication in proteins**. *Nat Struct Biol* 2003, **10**:59-69.