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).

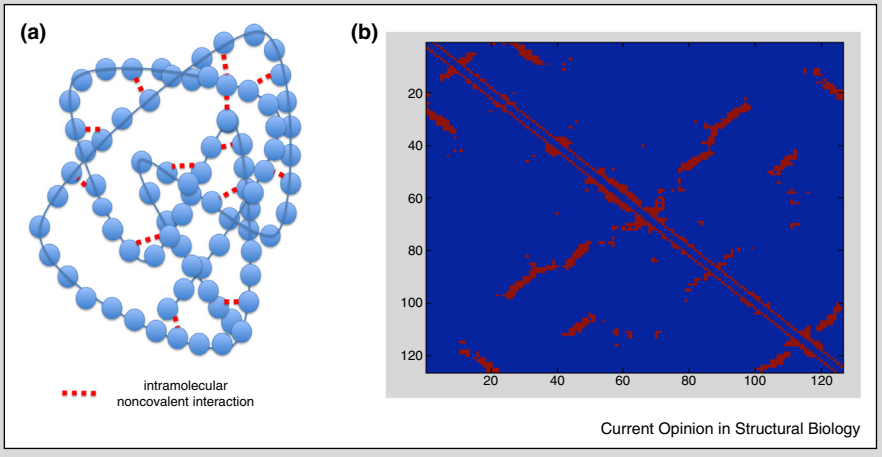
Figure 1 reports the modules (domains) identification in hemoglobin by spectral clustering (panel a), which identifies the four chains (consecutive blocks of different color, 141 residues in chains A and C, 146 in chains B and D). In addition, spectral decomposition highlights inter-module pathways, 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).