Biophysical Chemistry 165–166 (2012) 21–29

Contents lists available at SciVerse ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Shedding light on protein–ligand binding by graph theory: The topological nature of allostery

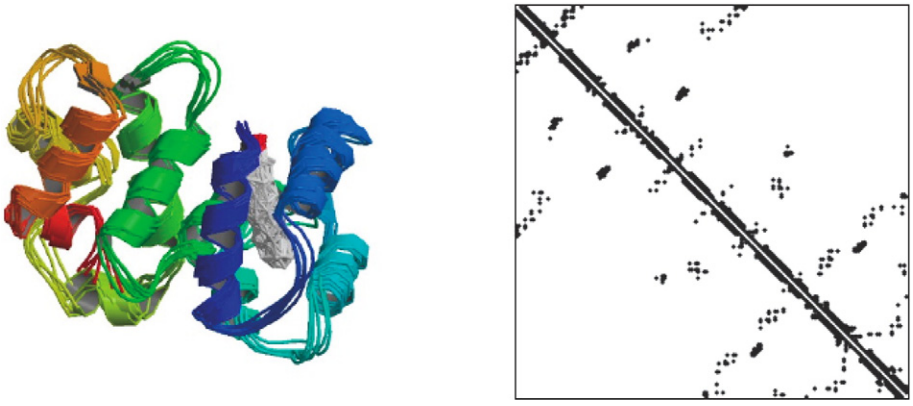
Micol De Ruvo [a], Alessandro Giuliani [b], Paola Paci [c], Daniele Santoni [d], Luisa Di Paola [a,*]

[a] Faculty of Engineering, Università CAMPUS BioMedico, Via A. del Portillo, 21, 00128 Roma, Italy
[b] Environment and Health Department, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161, Roma, Italy
[c] CNR-Institute of Systems Analysis and Computer Science (IASI), BioMathLab, viale Manzoni 30, 00185 Roma, Italy
[d] CNR-Institute of Systems Analysis and Computer Science (IASI), viale Manzoni 30, 00185 Roma, Italy

HIGHLIGHTS

- A graph theoretic method is proposed to analyze protein structure and function.
- The method is based on protein contact matrices.
- Application of the method shows the key role of topology in allosteric transitions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 February 2012
Received in revised form 2 March 2012
Accepted 2 March 2012
Available online 12 March 2012

Keywords:
Biological thermodynamics
Complex networks
Protein contact network
Enzyme kinetics
Bioinformatics

ABSTRACT

Allostery is a very important feature of proteins; we propose a mesoscopic approach to allosteric mechanisms elucidation, based on protein contact matrices. The application of graph theory methods to the characterization of the allosteric process and, more broadly, to obtain the conformational changes upon binding, reveals key features of the protein function. The proposed method highlights the leading role played by topological over geometrical changes in allosteric transitions. Topological invariants were able to discriminate between true allosteric motions and generic protein motions upon binding.

1. Introduction

Proteins are effective "biological operators": if genetic material can be considered as an information storage device, the effective biological work (whether catalysis, structure formation, signalling, immune defence, energy production..) is carried out by protein molecules and the need of their correct structure is crucial for life.

* Corresponding author. Tel.: +39 06225419634; fax: +39 0622541456.
E-mail address: l.dipaola@unicampus.it (L. Di Paola).

0301-4622/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2012.03.001

In the introduction of [1], Tanford and Reynolds discuss the general attitude of proteins to fulfill biological tasks: "For every imaginable task in a living organism, for every little step in every imaginable task, there is a protein designed to carry it out. The ultimate objective of a task may be chemical or mechanical or to measure colour or fight off a foreign invader — there is no limit to what can be accomplished".

This versatility of functions is due to a proper combination of structure stability and flexibility [2], that comes from several intramolecular interactions, differentially sensitive to the environmental stimuli. The 'correctness' of a protein structure comes from a delicate interplay between molecule and the chemico-physical microenvironment it is embedded into.

In this work, we assume a mesoscopic view of protein–environment interactions: protein structures are considered as networks, i.e. as graphs having the residues as nodes and their mutual contacts as edges, where aminoacid residues are indexed by a non-ambiguous ordering. This view has been demonstrated to be predictive in applications ranging from domain predictions [3] to allosteric hot spots [4].

A protein network corresponds to a square matrix having as rows and columns the aminoacid residues, ordered accordingly to their location along the sequence.

In the case of structural modelling, in the $\{i, j\}$ location we will have a no zero entry if there is a contact between the i-th and j-th residues in the 3D structure (see Materials and methods).

In Fig. 1.1 we report a simple protein structure (recoverin) and the corresponding residue adjacency matrix; the aminoacids ordering along the sequence makes the structure–adjacency matrix relation unique.

The contact network can be considered as the highway on which any signal (e.g., a structural modification) can be transmitted throughout the protein. These signals are the basis of protein physiological action. This creates an immediate and natural link between allostery and complex network analysis: the usual graph theoretical descriptors are, thus, ideal candidates to analyze protein structural and functional changes. Then, for instance, the degree of a node (i.e., the number of contacts of a specific residue in the protein structure) provides a natural description of the aminoacid relevance for protein structural stability (microscale, single residue, and level of modelling).

On the other hand, the spectral graph analysis of the protein network singles out specific modules of the protein (intermediate scale, clusters of the network). Eventually, different proteins can be compared in terms of their global graph descriptors, to elucidate similar construction principles (macroscale, entire protein level).

We focus on the characterization of the changes a protein structure undergoes when it binds its ligand. This characterization, to be effective, should be able to discriminate allosteric effects from mere ligand binding. It is well known that conformational changes are intrinsic to the function of a variety of proteins, and that are somehow triggered by ligand binding [5]. We analyze five proteins as for the differences between free (apo) and bound (holo) conformations. This comparison has been carried out by both geometrical and topological views, demonstrating the relevance of local topological description for the allosteric character identification.

2. Materials and methods

2.1. Protein systems

Calcium binding proteins (CaBPs) are responsible for many regulatory physiological mechanisms [6], switched by calcium concentration transients.

CaBP's can be divided into two general classes: sensors and buffers. Proteins belonging to the former class, like calmodulin (CaM) and recoverin, translate the chemical signal of an increased Ca^{2+} concentration into diverse biochemical responses. Calcium binding to sensor proteins leads to a transition from a tense (T) to a relaxed (R) state, that results in the exposure of a large hydrophobic surface, allowing the protein to interact with other molecular targets to accomplish regulatory functions.

On the other hand, the capture of calcium ions by buffer (or carrier) proteins, such as parvalbumin and calbindin D_{9k}, is accompanied by minor conformational changes. Sensors are allosteric protein, while buffers are not.

Our analysis has been extended to human hemoglobin and serum albumin, both representing a paradigm in protein science. Hemoglobin is in charge for the transport of respiratory gases and its structure was firstly resolved by Perutz, which was awarded of the Nobel Prize for this discovery [7]. Hemoglobin quaternary structure accounts for four-chains (A, B, C, D), each containing a heme group, responsible for oxygen and carbon monoxide binding, with a strong allosteric attitude [8].

Human serum albumin (HSA) is the most abundant serum protein, accounting for different biological functions, most of them linked to its ability to catch and carry hydrophobic molecules [9]. It is one of the largest single chain protein, showing different domains (I, II and III), each divided into two subdomains, A and B [9]. It has different binding sites for endogenous and exogenous toxins, drugs and metabolites [10]. The interaction between sites is limited to steric hindrance (steric interactions between fatty acids and L-tryptophan [11]) and it is not correlated to structural variations, neither at global nor at local scale. The non-allosteric character of HSA binding sites promotes its attitude as a carrier, making it able to transport independently several different classes of hydrophobic molecules, with large binding affinities.

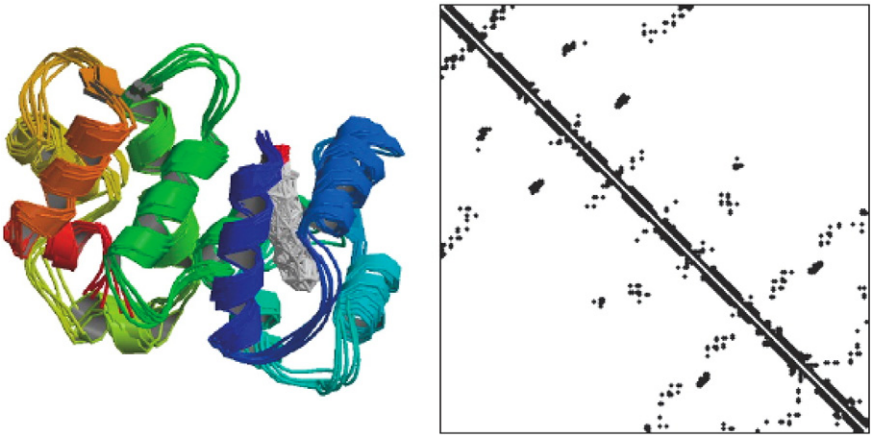


Fig. 1.1. Protein structure and adjacency matrix map for recoverin: contacts correspond to black dots.

2.2. Topological and geometrical comparison of protein structures

The protein structure was translated into a residue contact network, based on the Euclidean mutual distance matrix **d**, computed on the basis of the spatial positions of the α-carbons, extracted from the protein PDB file.

We established a cut-off for inter-residue distance ranging within $\mathcal{I} = [4-8]$ Å accounting for intramolecular noncovalent interactions [12]; thus, the corresponding unweighted protein structure graph was built up, whose adjacency matrix $A = \{a_{ij}\}$ is formally defined as:

$$a_{ij} = \begin{cases} 1 & \text{if } d_{ij} \in \mathcal{I} \\ 0 & \text{otherwise} \end{cases}$$

A key point is the identification of graph modules, namely groups of nodes showing a larger number of connections to each other with respect to those established with nodes outside the module itself. Such modules are strongly related to protein domains [13,14].

To single out meaningful modules, we applied the classical partitioning method based on the spectral analysis of the graph Laplacian matrix [15]. As a final result, the graph is divided into modules that are disjoint sets of nodes ("crispy" partition). Once the graph has been partitioned, for each node it is possible to evaluate two parameters that are a direct measure of its attitude to establish links with nodes belonging to the same module rather than with those lying into other modules [16]:

- the within-module z-score:

$$z_i = \frac{k_{is} - \bar{k}_{s_i}}{\sigma_{si}}$$

k_{is} is the number of links the i-th node establishes with nodes belonging to its own cluster s_i; $\bar{k}_{s_i}$ is the average degree for nodes in cluster s_i, and $LGRsv_{si}$ the corresponding standard deviation;

- the participation coefficient, that describes the attitude of the node to connect to other nodes off its own cluster:

$$P = 1 - \left(\frac{k_{is}}{k_i}\right)^2$$

k_i is the i-th node overall degree.

According to the values of P and z, it is possible to establish a cartography for the nodes [16], based on their role in terms of within- and between-modules connections.

High values of z combined with zero P are characteristic for nodes occupying a central position in their own module, while not even communicating with other modules; on the other way, nodes that are characterized by not-null P values make modules communicate with each other (signal transmission).

We make the hypothesis this latter process is the molecular engine behind the allosteric effect. The demonstration of a hypothetical percolation threshold of allosteric proteins in terms of such a description will constitute a further proof of our statement.

To investigate structural similarity between the apo and holo forms, we computed the Root Mean Square Distance (*RMSD*) [17] at a global and a local scale.

As we will discover in the next section, an active site residue going from a not-null to a zero value of P corresponds to a shift from a "prone to binding" to a "closed" state. Eventually, we introduced another topological parameter, the Hamming distance D between adjacency lists, that reports the overall number of matrix positions at which the corresponding values are different.

In details, two metrics have been computed: the first one (D_{av}) is the Hamming distance normalized to the corresponding connectivity of a completely connected graph ($\frac{N(N-1)}{2}$), while the other (D_{eff}) is normalized to the number of apo form effective contacts (unit values in the apo adjacency matrix) .

Global (entire protein) and local (single residue) metrics are summarized in Table 2.

Table 1
List of proteins mentioned with reference to the specific name, PDB ID, family and class (the classification is based on the relative ligand binding function).

Protein name	PDB ID	
	Apo	Holo
EF-hand calcium binding proteins		
Calbindin D_{9k} (Cb-D_{9k}) (buffer)	1CLB	2BCA
Parvalbumin (PV) (buffer)	2NLN	1TTX
Recoverin (RC) (sensor)	1IKU	1JSA
Human hemoglobin (HbO2, HbCO)	2DN2	1GZX (O2), 1BBB (CO)
Human serum albumin (Ab)	1AO6	2VUE (bilirubin)

Table 2
Global and local parameters employed for topological and structural analysis.

Global metrics		Local metrics	
Topological	Structural	Topological	Structural
D_{av}	RMSD	*P,z* (active sites)	*RMSD* per residue
D_{eff}			

3. Results

3.1. Global level analysis

At first, we computed the topological distances D_{av} and D_{eff} between the apo and holo forms for proteins in Table 1; results reported in Table 3 show that D_{av} scales with protein size: this is a logical consequence of the exponential decay of contacts with size [13,18].

On the other hand, D_{eff} does not appear related neither to the size nor to allostery, being strongly invariant across different protein systems, despite their size, kinetic properties and, as we will discover, huge differences in flexibility (measured by *RMSD*).

This suggests the contact matrix maintains the protein identity and mirrors the strong invariance we observed as for *P–z* global distributions of proteins [13].

On the other hand, results in Table 4 show a huge variation in *RMSD* values regardless of size or allosteric character. Thus, global *RMSD* value does not provide any useful information about allosteric effect.

3.2. Local level

Hemoglobin is assimilable to sensor proteins, it undergoes conformational changes due to its physiological activity (respiratory gases binding). The quaternary configuration of low affinity, deoxygenated hemoglobin is known as the tense (T) state, whereas the quaternary structure of the fully oxygenated high affinity form is known as the relaxed (R) state. The dissociation constant of the first oxygenation step is $K_T = 1 \cdot 10^{-4}$ M while that related to the last oxygenation step high affinity) is $K_R = 2 \cdot 10^{-6}$ M [19].

Table 3
D_{av} and D_{eff} for sample proteins.

	Cb-D_{9k}	PV	RC	HbO2	HbCO	Ab
N	75	108	188	574	574	1124
D_{av}	0.015	0.012	0.006	0.002	0.002	0.001
D_{eff}	0.58	0.62	0.53	0.56	0.60	0.56

Table 4
Global *RMSD* calculated for sample proteins.

	Cb-D_{9k}	PV	RC	HbO2	HbCO	Ab
N	75	108	188	574	574	1124
RMSD	1.84	3.59	11.45	0.30	1.21	0.94

When carbon monoxide is present, it competes strongly with oxygen at the heme binding sites (the dissociation constant is 200 times smaller than K_R), forming a very bright red form of hemoglobin called carboxyhemoglobin; the corresponding dissociation constants are $K_T = 1 \cdot 10^{-6}$ M and $K_R = 5 \cdot 7^{-8}$ M.

The results of *P–z* analysis of the deoxigenated (2DN2), oxygenated (1GZX) and carbon-monoxy (1BBB) forms, are shown in Table 5, where A–C and B–D chains are grouped, due to the symmetry of hemoglobin structure. It appears clear that in the holo states, *P* values in the active sites turn to 0 (this implies that the active site is decoupled from other modules). We interpreted this shift in terms of a T–R transition, coupled with an increase of ligand binding affinity. Also for the *CO*-bound complex all *P* values in the binding site turn to 0, confirming the higher affinity of hemoglobin for this ligand.

We have also reported the index ($\overline{P}_0$) that expresses the fraction of residues having $P = 0$, for each protein form. The net increase of null values in the *P* distribution in holo forms with respect to the apo one is very remarkable, given the basic invariance of *P–z* distribution, and it is consistent with the "closing of the gates" between modules, hindering the signal transmission throughout the protein.

HSA binds bilirubin, a toxic metabolite of heme (dissociation constant: $K_d = 10^{-7}$ M–10^{-8} M) at a high affinity site and acts as a buffer preventing the transfer of bilirubin from blood to tissues that otherwise would cause bilirubin encephalopathy [20]. Focusing our analysis on contacts belonging to this site the non-allosteric character of HSA domains corresponds to *P* values showing only minor changes upon binding, as reported in Table 6.

3.3. Calcium binding proteins

Recoverin is a calcium-binding protein involved in visual signal transduction; the protein molecule is composed of two domains, each bearing the calcium binding active sites [21]. Recoverin undergoes a typical T (tense)–R (relaxed) structural transition upon calcium binding [22]; the dissociation constants of the two sites in the holo state are $K_R = 1 \cdot 4^{-5}$ M for site 1 and $K_T = 9.1 \cdot 10^{-6}$ M for site 2, respectively. As we can observe in Table 7, our data confirm the higher affinity of site 2: all residues of this site have $P = 0$ in the holo state. Moreover, on the entire protein, the $P = 0$ nodes increase in number in the holo form, with respect to the apo one (Table 7).

Parvalbumins (PVs) are a family of small, acidic Ca^{2+} binding proteins found in muscle of vertebrates, in mammal brain and endocrine glands [23]. Binding and release of Ca^{2+} induce relatively small conformational changes [24], confined within the Ca^{2+} binding loop or in the close vicinity. Therefore, the main function of PV is presumably the control and modulation of intracellular Ca^{2+} signals [25].

Table 5
Hemoglobin *P* values for the active sites aminoacids relative to the deoxygenated (2DN2), complexed with oxygen (1GZX) and with carbon-monoxy (1BBB) forms.

	Chain A-C					Chain B-D				$\overline{P}_0$
	Leu	Phe	His	Val	Leu	Leu	Phe	His	Val	
2DN2 (apo)	0	0.30	0	0.43	0.39	0	0	0	0.26	0.13
1GZX (holo)	0	0.31	0	0	0	0	0	0	0	0.70
1BBB (holo)	0	0	0	0	0	0	0	0	0	0.75

Table 6
Apo (1AO6) and holo (2VUE) HSA *P* values for the active sites aminoacids.

Active site	1AO6 (apo)	2VUE (holo)
Arg	0.96	0.64
Leu	0.96	0.43
Val	0.98	0.75
Arg	0.84	0.67
Pro	0.75	0.94
Val	0.86	0.61
Met	0.86	0.67
Ala	0.75	0.51
Phe	0.82	0.89
Lys	0.75	0.89
Tyr	0.75	0.95
Glu	0.19	0.69
Ile	0.61	0.89
Arg	0.91	0.75
His	0.99	0.75
Phe	0.86	0.75
Leu	0.56	0.89
Phe	0.86	0.75
Tyr	0.69	0.75
Leu	0.75	0.56
Leu	0.67	0.86
Arg	0.75	0.84
Gly	0.51	0.91
Lys	0.89	0.75
$\overline{P}_0$	0.013	0.0125

Rat parvalbumin contains two equivalent Ca^{2+} binding sites, (dissociation constant $K_d = 11$nM) binding CA^{2+}in a noncooperative way [23].

In this study, apo and holo form of rat LGRb-parvalbumin are considered for the analysis: the two domains bind Ca^{2+} with similar affinities ($K_1 \approx K_2 = 4 \cdot 10^{-8}$ M) and, in this case too, alterations in *P* values reflect thermodynamic data; actually, although the number of $P = 0$ nodes does not increase in the holo state (there is no global conformational change), in the active site, either in site 1 and 2, the residues *P* value goes to 0 (Table 8).

We can conclude that, even if a local change occurs, binding sites affinities are not modified, as expected for a non-allosteric binding mechanism.

Calbindin D_{9k} (Cb-D_{9k}) is the smallest EF-hand protein, whose EF-hand pair consists of a "canonical" EF-hand domain with a Ca^{2+} binding loop (EF2) and a non-canonical (also termed "pseudo EF-hand") Ca^{2+} binding loop (EF1), typical for the S100 family proteins. The two EF-hand domains bind Ca^{2+} with similar affinities ($K_1 \approx K_2 = 4 \cdot 10^{-8}$ M) and positive cooperativity; it has been seen that there are long range effects that contribute to the positive cooperativity of calcium binding to calbindin. In Johnson et al.'s [26] study, the flexibility of the methyl groups of the apo versus the holo state has been examined: although the overall Ca^{2+} induced conformational change in Cb-D_{9k} is not pronounced, upon binding, the methyl groups (and thus the rest of the protein bound to them) at the end of the protein far from the binding sites becomes more flexible [27]. This behaviour is mirrored by the change in *P* values distribution depicted in Fig. 3.1.

Thus, we can again prove the consistency of our hypothesis: as it can be observed in Table 9, not only $\overline{P}_0$ values do not change, but even *P* values of the active sites remain substantially unchanged, pointing to the absence of evident modification in the topological structure of the whole protein.

Table 7
Apo (1IKU) and holo (1JSA) recoverin *P* values for the active sites aminoacids.

	Site 1					Site 2				$\overline{P}_0$
	Asp	Asn	Asp	Thr	Glu	Asp	Asn	Thr	Glu	
1IKU (apo)	0	0	0	0	0.67	0.19	0	0	0	0.54
1JSA (holo)	0.56	0	0.36	0.26	0	0	0	0	0	0.56

Table 8
Apo (2NLN) and holo (1TTX) parvalbumin *P* values for the active sites aminoacids.

	Site 1					Site 2				$\overline{P}_0$
	Asp	Ser	Tyr	Glu	Asp	Asp	Asp	Lys	Glu	
2NLN (apo)	0.61	0	0	0	0	0	0	0.16	0.75	0.41
1TTX (holo)	0	0	0	0	0	0	0	0	0	0.37

Table 9
Apo (1CLB) and holo (2BCA) calbindin D_{9k} *P* values for the active sites aminoacids.

	Site 1					Site 2				$\overline{P}_0$
	Ala	Glu	Asp	Gln	Glu	Asp	Gly	Glu	Glu	
1CLB (apo)	0	0	0	0.51	0.75	0.19	0.61	0.75	0.75	0.49
2BCA (holo)	0	0	0	0.36	0.19	0.17	0.36	0.69	0.21	0.39

Therefore, the allosteric mode (recoverin) amplifies the local change (in the active site) on a global scale rearrangement; on the contrary, the non-allosteric mode (parvalbumin) corresponds only to local scale topology variations (see Fig. 3.2(a)).

As depicted in Fig. 3.2(b), in the case of recoverin the distribution changes throughout the whole protein, while in the case of parvalbumin the modifications are confined to regions close to the active site.

Eventually, we computed *RMSD* for residues in the active site of proteins (Fig. 3.3): these values are quite smaller than global *RMSD*. This result, together with the lack of discrimination ability of RMSD as for allosteric/non-allosteric systems, is a further proof of the mainly topological nature of allosteric motion, while geometrical changes are common to any kind of protein–ligand binding.

4. Discussion

It is important to put our results into the general perspective of protein–ligand interactions, that have been largely examined in terms of ligand binding thermodynamics equilibrium [28,29]. Simple models refer to the very simple case of n active sites on a protein P, that are independent and identical (same microscopical dissociation constant k); in this case, the binding equilibrium is represented by the well-known Scatchard equation:

$$\frac{\nu}{c_L} = \frac{n}{k} - \frac{c_L}{k} \tag{4.1}$$

where c_L is the free ligand concentration and ν represents the equilibrium saturation degree, defined as:

$$\nu = \frac{\sum_{i=1}^{n} i \cdot c_{Pi}}{c_{P0}} \tag{4.2}$$

c_{Pi} is the ligand–protein complex (accounting for i molecules of L bound to the protein active sites, such that the overall protein concentration is $c_{P0} = c_P + \sum_{i=1}^{n} c_{Pi}$).

Applying the mass balance to all molecular species it is derived an expression in terms of directly observable variables (c_L, c_{P0} and c_{L0}):

$$\nu = \frac{c_{L0} - c_L}{c_{P0}} \tag{4.3}$$

where c_{L0} is the overall ligand concentration.

Once n is known, the corresponding fraction saturation degree y as function of c_L is described by the typical saturation hyperbolic curve:

$$y = \frac{\sum_{i=1}^{n} i \cdot c_{Pi}}{n \cdot c_{P0}} = \frac{\nu}{n} = \frac{c_L/k}{1 + c_L/k} \tag{4.4}$$

This representation is often oversimplified: frequently, the ligand interaction with the active site leads to a local and, more often, to a global conformational transition, thus changing the ligand affinity of the remaining free active sites. This phenomenon is addressed to as allostery and it results in a dissociation constant k varying with the saturation degree ν.

Allostery is a specific aspect of cooperativity, underlying many biochemical processes such as folding [30], biomacromolecules aggregation [31] and, indeed, biomacromolecules ligand binding.

Cooperativity derives from the complex chemico-physical nature of biomacromolecules, giving rise to a continuous exchange between local and global — over the all molecule — events [32]. Thus, the result of multiple local processes is almost always not additive with respect to the single process contribution; the difference is a measure of the ability of the local events to be propagated on a larger scale.

Let's set k_0 as the value of k at zero saturation degree, being $\Delta G_0^0 = -RTlog(k \cdot c_{rs})$ the corresponding standard specific Gibbs free energy (c_{rs} is the standard reference for the concentration, taken 1 M), the value of the dissociation Gibbs free energy at the saturation degree ν can be expressed as [28]:

$$\Delta G_\nu^0 = \Delta G_0^0 + RT\ \phi(\nu) \tag{4.5}$$

$\phi(\nu)$ is a measure of the site interaction energy, at a given saturation degree ν.

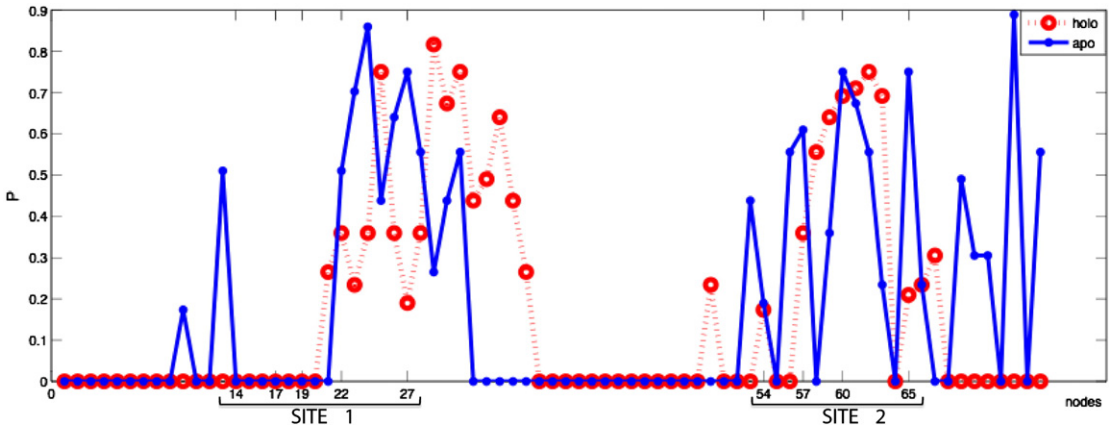


Fig. 3.1. *P* values distribution for apo (blue) and holo form (red) of calbindin; residues involved in the active sites are emphasized. Flexibility of the holo form is evident as *P* values of several residues out of the active sites turn into 0.

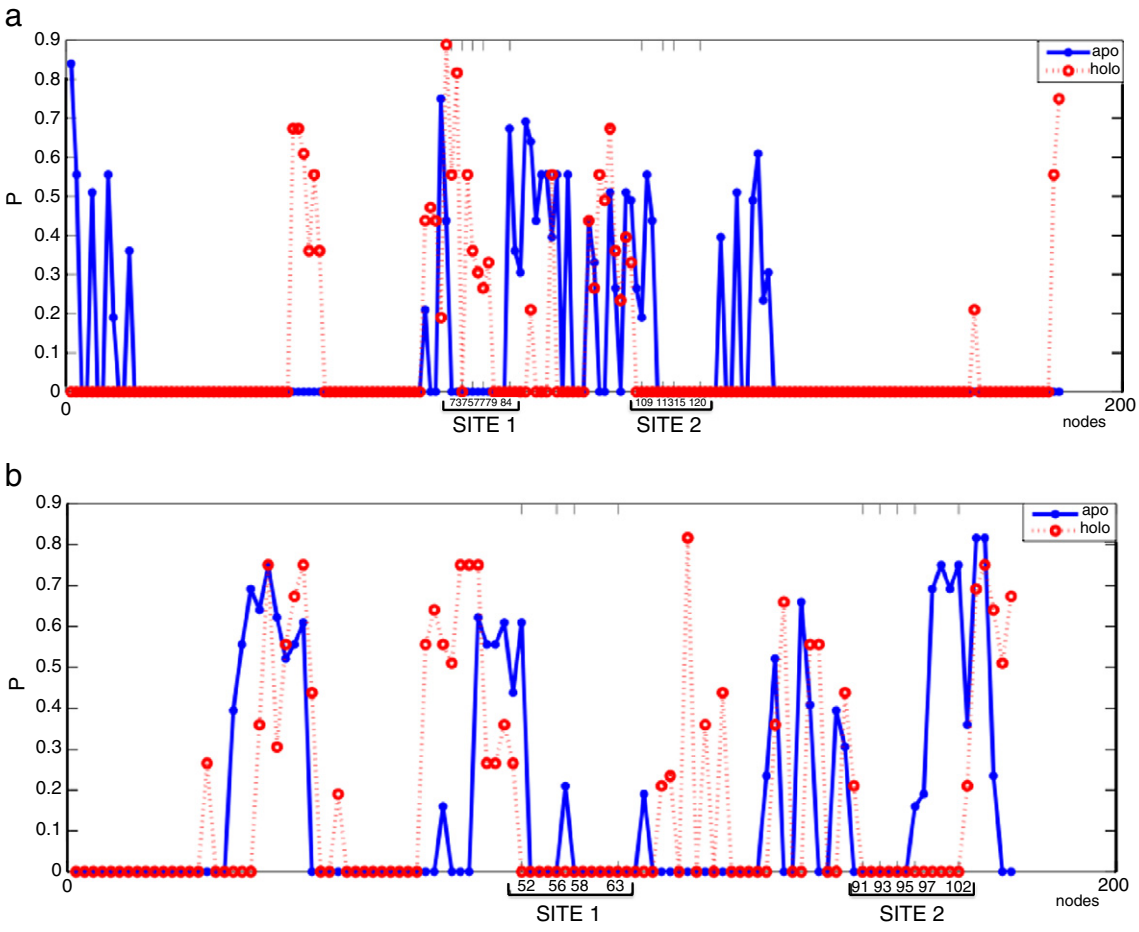


Fig. 3.2. *P* values distribution for apo (blue) and holo form (red) of recoverin a) and of parvalbumin b). Residues involved in the active sites are emphasized. Overall *P* values sharply change in the holo form of RC whereas, in the case of PV, holo form P distribution quite follows that of the apo one.

Since $\Delta G_\nu^0 = -RT\,logk(\nu)$, it is obtained:

$$k(\nu) = k_0 \cdot exp(-\phi(\nu)) \tag{4.6}$$

Unfortunately, it is not straight and simple to derive a physically-based description of $\phi(\nu)$; in any case, when k depends on ν, if all the residue sites have initial identical affinity, an allosteric process is probably occurring.

A semiempirical approach [33] is based on an "all-or-none" scheme:

$$P + \alpha L \rightleftharpoons P \tag{4.7}$$

to which corresponds an apparent macroscopic dissociation constant

$$K_\alpha = \frac{c_P \cdot c_L^\alpha}{c_{P\alpha}} \tag{4.8}$$

the corresponding fractional saturation degree y is:

$$y = \frac{(c_L/K)^\alpha}{1 + (c_L/K)^\alpha} \tag{4.9}$$

where the parameter α, known as the Hill constant, is derived as:

$$\alpha = \frac{d\left\{log\left[\frac{y}{1-y}\right]\right\}}{d[log(L)]} \tag{4.10}$$

α lies in the range $[1, n]$ and it is a direct measure of the cooperativity of the binding process: when $\alpha = 1$, Eq. (4.10) equals to Eq. (4.4), describing a noncooperative system; on the other hand, when $\alpha = n$, the process is named as infinitively cooperative, indicating a simultaneous binding of ligand molecules to all protein active sites.

Classical molecular models for allosterism are based on the concept that interactions between sites is translated into conformational perturbations that locally modify binding site shape and, consequently, ligand affinity; this conformational change is thought to act also at a global scale, giving rise to different structural forms, whose transformation equilibria match to those of binding [28].

If the identical binding sites are located on highly similar chains that are collected together into a quaternary structure, the MWC model [34] describes the concerted motion of subunits due to ligand binding, that results into an allosteric behaviour; this model has been applied to describe the affinity of hemoglobin for respiratory gases (oxygen and carbon dioxide); if more complex conformational changes are involved, the KNF model [35] is able to describe the relationship between sites occupancy and affinities.

These idealized representations are based on a vision of the proteins, as a rigid objects that change mostly abruptly their shape from one conformational state to another. Recently, the linkage between conformational changes and allosterism has been thoroughly discussed [36]: the allosteric conformational change may be revised in terms of redistribution of conformation distribution that assets the native structure conformation ensemble. In other words, allosterism doesn't induce conformational changes to the native, apoprotein form, but it shifts the conformational state distribution for the apo form towards the conformations corresponding to the higher ligand affinities (R state, for the MWC model, for instance) for cooperative allosterism. In other words,

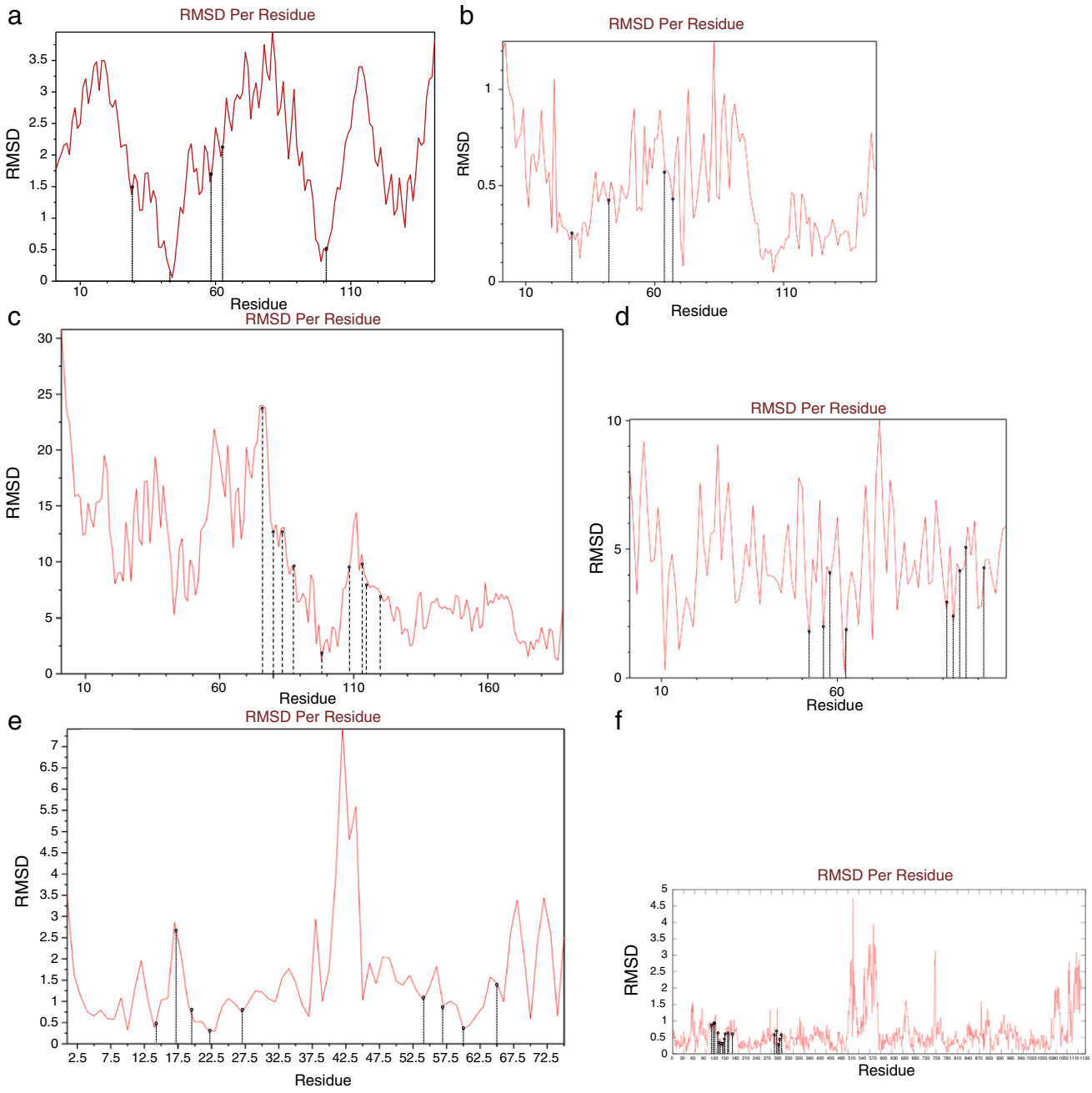


Fig. 3.3. *RMSD* per residue resulting from the superposition of the apo and holo forms of proteins: a) chain A and b) chain B of 2DN2-1GZX (Hb); c) 1IKU-1JSA (RC); d) 1TTX-2NLN (PV); e) 2BCA-1CLB (Cb-D_{9k}); f) 1AO6-2VUE (Ab). Dashed lines are traced to highlight active sites residues: for each protein, peaks correspondent to these residues are not prominent with respect to the global *RMSD* tendency.

the ligand binding doesn't create *de novo* conformations that are not energetically allowed in the apo form, but simply increase their probability of existence. Another point of view involves the dynamic fluctuations of protein atoms, that are strongly related to the intramolecular protein structure interactions, strongly influenced by the protein microenvironment [37–39]. Protein dynamics describe how proteins breathe in their environment, acting and reacting to external stimuli; to measure the entity of local and global fluctuations, the Root Mean Square Distance (*RMSD*) is successfully applied [40]. What is relevant for our approach is that, besides the different models of allosterism, a fundamental kinetic difference linked to the value of parameter α does exist between allosteric and not allosteric systems. Thus, it could be in principle highlighted by a topological/structural analysis.

The definition of what is allosteric and what is not, in our opinion, must be based on reaction kinetics; on the contrary, if we assume a purely structural definition of "a motion started at a given site and further generalized to other parts of the systems", each protein system can be considered as allosteric, as correctly stressed by Del Sol and Nussinov [4].

Applying our method, we were able to identify global motions discriminating apo and holo structures in all the studied systems (both allosteric and not allosteric).

The analysis presented demonstrates how topology is useful to discriminate proteins structure variation upon ligand binding, through the explanation of contact roles. Actually, our results indicate that the variation in *P* values is a peculiar hint to discriminate allosteric behaviour of sensor proteins despite to the rigid structure of the buffer ones. The negligible (but strongly invariant) topological motion of aminoacids does not provoke relevant global changes in contact network but, on a local scale, it leads to the variation in the role of contacts themselves;

meaning, the loss of boundary contacts within clusters being spreaded, let the entire structure being free to move. Thus, it can be easily understood that clusters individuated by spectral theory have a functional role.

5. Conclusion

We proposed a simple methodology, based on a network interpretation of protein structures, and applied it to the analysis of CaBP's, a well-known class of proteins involved in calcium metabolism by different molecular mechanisms, giving a proof-of-concept of the possibility to link chemico-physical features to protein graph invariants.

It can be therefore stated that the variation in P values is a peculiar hint to discriminate allosteric behaviour of sensor proteins despite to the rigid structure of the buffer ones.

We give a further proof of how graph theory provides a useful general framework for proteins science.

Future directions of our method could lead to the detection of altered values in the network modules, and/or in the *P–z* distribution of the aminoacids, thus identifying clusters that are pathologically altered in diseases, like Alzheimer's and Parkinson's. [25]

Furthermore, the identification of network invariants able to follow the variation of structure due to binding, in the physiological and pathological scenario, could be a starting point for drug design: the simulation *in silico* of synthetic CaBP's could be guided through the forecast of affinity for the active site through the network analysis of novel apo and holo forms.

The nature of global motions induced in a protein system by the binding of a ligand was investigated by means of molecular dynamics approaches by [41,42] , pointing to the very organized character of such motions. The conformational changes can be described as motion along one or few coherent degrees of freedom: only pairs of sites that couple to the same conformational degrees of freedom (e.g. sharing the same loading profile along the principal components of motions [43,44]) can be allosterically connected [41]. This 'coupling of sites' is the dynamic analogue to our static 'between modules' interaction as modeled by P index [43,44] . This analogy is made more cogent by the results of Park and Kim [45] which find a correlation between 'allosteric hotspots' (residues more deeply involved in allosteric signal transmission) and the 'betweeness' of the corresponding nodes in the protein contact network (a measure linked to the number of different pathways passing by a given node). The general consistency between dynamical (correlation between motions of residue pairs) and structural (distance in space between residue pairs) protein representations was given a proof-of-concept in [46], so in some sense 'closing the circle' between dynamical and topological allosteric behaviour pictures.

Our results confirm the hypothesis the allosteric signal transmission relies on non covalent contacts with only a minor (if any) role exerted by the peptide bonds backbone [47–49,4]. Indeed, topological metrics prevailed over geometrical one with reference to allosteric and non-allosteric discrimination. In this respect, it is worth noting that the 'global motion' is common to any protein system in the passage from apo to holo state. This is a consequence of the fact the protein molecule is a strongly connected entity.

On the other hand, geometrical and topological global motions have a very different status: while geometrical motions can vary a lot between different systems due to the relative flexibility of the structure, topological motions are strongly invariant across very different molecules. This allows us to hypothesize the 'contact network' as the invariant core responsible for maintaining protein identity, so allowing only relatively minor (and strongly concerted) motions. The fact topological changes are highly concerted implies the 'quality' of motions is more important than their mere entity. Allosteric 'effective' motions are of a very special kind in terms of their effects on protein networks: they 'close' the communication lines between active site and the rest of the molecule, affecting the inter-modules pathways. This behaviour was not observed in non-allosteric system and stresses the importance of protein network description in studying structure-function relations. Our result open a possible avenue so to rationalize the search for the so called 'allosteric-network' drugs devised by Csermely and Nussinov [50], in order to be really effective, our approach should be enlarged to natively-unfolded system (e.g. by the use of NMR data), given their relevance in cell signaling [51,52]. A last remark is linked to the need of assuming a thermodynamic based definition of allostery being the simple statement 'motion transmitted at distance from the ligand' being common to any substrate binding event.

References

[1] C. Tanford, J. Reynolds, Nature's Robots: a History of Proteins, Oxford University Press, 2004.
[2] K. Teilum, J.G. Olsen, B.B. Kragelund, Protein stability, flexibility and function, Biochimica et Biophysica Acta 1814 (8) (2011) 969–976.
[3] S. Vishveshwara, K. Brinda, N. Kannan, Protein structure: insights from graph theory, Journal of Theoretical and Computational Chemistry 1 (1) (2002) 187–212.
[4] A. del Sol, M. Araúzo-Bravo, D. Amoros, R. Nussinov, Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages, Genome Biology 8 (5) (2007) R92.
[5] F. Capozzi, F. Casadei, C. Luchinat, Ef-hand protein dynamics and evolution of calcium signal transduction: an NMR view, Journal of Biological Inorganic Chemistry 11 (8) (2006) 949–962.
[6] E. Permyakov, R. Kretsinger, Calcium Binding Proteins, Wiley Series in Protein and Peptide Science, John Wiley and Sons, New York-London, 2011.
[7] M. Perutz, M. Rossmann, A. Cullis, H. Muirhead, G. Will, A. North, Structure of hemoglobin, Nature 185 (4711) (1960) 416–422.
[8] J. Edsall, Hemoglobin and the origins of the concept of allosterism, Federation Proceedings 39 (2) (1980) 226–235.
[9] T.P. Jr, Serum albumin, Advances in Protein Chemistry 37 (1975) 161–245.
[10] X. He, D. Carter, Atomic structure and chemistry of human serum albumin, Nature 358 (1992) 209–215.
[11] V. Cunningham, L. Hay, H. Stoner, The binding of l-tryptophan to serum albumins in the presence of non-esterified fatty acids, Biochemical Journal 146 (1975) 653–658.
[12] I. Bahar, R. Jernigan, Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation, Journal of Molecular Biology 266 (1) (1997) 195–214.
[13] A. Krishnan, A. Giuliani, J. Zbilut, M. Tomita, Network scaling invariants help to elucidate basic topological principles of proteins, Journal of Proteome Research 6 (10) (2007) 3924–3934.
[14] A. Giuliani, L.D. Paola, R. Setola, Proteins as networks: a mesoscopic approach using haemoglobin molecule as case study, Current Proteomics 6 (4) (2009) 235–245.
[15] M. Newman, Finding community structure in networks using the eigenvectors of matrices, Physical Review E 74 (3) (2006) 036104.
[16] R. Guimerà, L. Amaral, Cartography of complex networks: modules and universal roles, Journal of Statistical Mechanics P02001 (2005) 1–13.
[17] O. Carugo, S. Pongor, A normalized root-mean-square distance for comparing protein three-dimensional structures, Protein Science 10 (7) (2001) 1470–1473.
[18] L. Di Paola, P. Paci, D. Santoni, M. De Ruvo, A. Giuliani, Proteins as sponges: a statistical journey along protein structure organization principles, Journal of Chemical Information and Modeling 52 (2) (2012) 474–482.
[19] K. Imai, Adair fitting to oxygen equilibrium curves of hemoglobin, Methods in Enzymology 232 (1994) 559–576.
[20] C. Petersen, C. Ha, K. Harohalli, J. Feix, N. Bhagavan, A dynamic model for bilirubin binding to human serum albumin, The Journal of Biological Chemistry 275 (28) (2000) 20985–20995.
[21] S. Permayakov, A. Cherskaya, I. Senin, A. Zargarov, S. Shulga-Morskoy, A. Alekseev, D. Zinchenko, V. Lipkin, P. Philippov, V. Uversky, E. Permyakov, Effects of mutations in the calcium-binding sites of recoverin on its calcium affinity: evidence for successive filling of calcium binding sites, Protein Engineering 13 (11) (2000) 783–790.
[22] J. Ames, T. Porumb, T. Tanaka, M. Ikura, L. Stryer, Amino-terminal myristoylation induces cooperative calcium binding to recoverin, The Journal of Biological Chemistry 270 (9) (1995) 4526–4533.
[23] M. Eberhard, P. Erne, Calcium and magnesium binding to rat parvalbumin, European Journal of Biochemistry 222 (1994) 21–26.
[24] W. Yang, H. Lee, H. Hellinga, J. Yang, Structural analysis, identification, and design of calcium-binding sites in proteins, Proteins 47 (3) (2002) 344–356.
[25] B. Schwaller, The continuing disappearance of "pure" Ca2 + buffers, Cellular and Molecular Life Sciences 66 (2) (2009) 275–300.
[26] S. Fischer, K. Olsen, K. Nam, M. Karplus, Unsuspected pathway of the allosteric transition in hemoglobin, PNAS 108 (14) (2011) 5608–5613.
[27] C. Bode, Network analysis of protein dynamics, FEBS Letters 581 (2007) 2776–2782.
[28] C. Cantor, P. Schimmel, The behaviour of biological macromolecules, WH Freeman and Company, New York, 1971.

[29] C. Tanford, Physical Chemistry of Macromolecules, John Wiley and Sons, New York-London, 1961.
[30] K. Dill, K. Fiebig, H. Chan, Cooperativity in protein-folding kinetics, PNAS 90 (1993) 1942–1946.
[31] D. Hamada, T. Tanaka, G. Tartaglia, A. Pawar, M. Vendruscolo, M. Kawamur, A. Tamura, N. Tanaka, C. Dobson, Competition between folding, native-state dimerisation and amyloid aggregation in beta-lactoglobulin, Journal of Molecular Biology 386 (2009) 878–890.
[32] E. Di Cera (Ed.), Thermodynamics in Biology, Oxford University Press, 2000.
[33] A. Hill, The combinations of haemoglobin with oxygen and with carbon monoxide, The Journal of Physiology 40 (1910) 4–7.
[34] J. Monod, J. Wyman, J. Changeux, On the nature of allosteric transitions: a plausible model, Journal of Molecular Biology 12 (1965) 88–118.
[35] D. Koshland, G. Nemethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1966) 365–385.
[36] C. Tsai, A. del Sol, R. Nussinov, Allostery: absence of a change in shape does not imply that allostery is not at play, Journal of Molecular Biology 378 (1) (2008) 1–11.
[37] N. Goodey, S. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nature Chemical Biology 4 (8) (2008) 474–482.
[38] M. Lee, J. Tsai, D. Baker, P. Kollman, Molecular dynamics in the endgame of protein structure prediction, Journal of Molecular Biology 313 (2001) 417–430.
[39] J. Wand, Dynamic activation of protein function: a view emerging from NMR spectroscopy, Nature Structural & Molecular Biology 8 (2001) 926–931.
[40] P. Steinbach, B. Brooks, Protein hydration elucidated by molecular dynamics simulation, PNAS 90 (1993) 9135–9139.
[41] S. Mitternacht, I.N. Berezovsky, Coherent conformational degrees of freedom as a structural basis for allosteric communication, PLoS Computational Biology 7 (12) (2011) e1002301.
[42] Z.N. Gerek, S.B. Ozkan, Change in allosteric network affects binding affinities of pdz domains: analysis through perturbation response scanning, PLoS Computational Biology 7 (10) (2011) e1002154.
[43] M. Bhattacharyya, A. Ghosh, P. Hansia, S. Vishveshwara, Allostery and conformational free energy changes in human tryptophanyl-trna synthetase from essential dynamics and structure networks, Proteins Structure & Function Bioinformatics 78 (3) (2010) 506–517.
[44] M. Bhattacharyya, S. Vishveshwara, Probing the allosteric mechanism in pyrrolysyl-trna synthetase using energy-weighted network formalism, Biochemistry 50 (28) (2011) 6225–6236.
[45] D.K. Keunwan Park, Modeling allosteric signal propagation using protein structure networks, BMC Bioinformatics 12 (Suppl. 1) (2011) 1471–2105.
[46] A. Kloczkowski, R. Jernigan, Z. Wu, G. Song, L. Yang, A. Kolinski, P. Pokarowski, Distance matrix-based approach to protein structure prediction, Journal of Structural and Functional Genomics 10 (1) (2009) 67–81.
[47] S. Lockless, R. Ranganathan, Evolutionarily conserved pathways of energetic connectivity in protein families, Science 286 (5438) (1999) 295–299.
[48] G. Suel, S.W. Lockless, M. Wail, R. Ranganathan, Evolutionarily conserved networks of residues mediate allosteric communication in proteins, Nature Structural & Molecular Biology 10 (1) (2002) 59–69.
[49] A. del Sol, H. Fujihashi, D. Amoros, R. Nussinov, Residues crucial for maintaining short paths in network communication mediate signaling in proteins, Molecular Systems Biology 2 (2006) 0019.
[50] R. Nussinov, C. Tsai, P. Csermely, Allo-network drugs: harnessing allostery in cellular networks, Trends in Pharmacological Sciences 32 (2011) 686–693.
[51] A. Dunker, I. Silman, V. Uversky, J. Sussman, Function and structure of inherently disordered proteins, Current Opinion in Structural Biology 18 (2008) 756–764.
[52] V. Uversky, Protein folding revisited. a polypeptide chain at the folding–misfolding–nonfolding cross-roads: which way to go? Cellular and Molecular Life Sciences 60 (9) (2003) 1852–1871.