Journal of Molecular Graphics and Modelling 40 (2013) 30–39

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

# Toward an understanding of the sequence and structural basis of allosteric proteins

Xiaobai Li[a,1], Yingyi Chen[a,b,1], Shaoyong Lu[a,b,1], Zhimin Huang[a,b], Xinyi Liu[a,b], Qi Wang[a], Ting Shi[a,b], Jian Zhang[a,b,c,*]

[a] Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai JiaoTong University, School of Medicine, Shanghai 200025, China
[b] Medicinal Bioinformatics Center, Shanghai JiaoTong University, School of Medicine, Shanghai 200025, China
[c] Shanghai Key Laboratory for Tumor Microenvironment and Inflammation, Shanghai JiaoTong University, School of Medicine, Shanghai 200025, China

ARTICLE INFO

*Article history:*
Accepted 19 December 2012
Available online 5 January 2013

*Keywords:*
Allostery
Allosteric protein
Allosteric site
Orthosteric site

ABSTRACT

Allostery is the most efficient means of regulating protein functions, ranging from the control of metabolic mechanisms to signal transduction pathways. Although allosteric regulation has been recognized for half a century, our knowledge is limited to the characteristics of allosteric proteins and the structural coupling of allosteric sites and modulators. In this paper, we present a comprehensive analysis of allosteric proteins that provides insight into the foundation of allosteric interactions by revealing a series of common features in the allosteric proteins. Allosteric proteins mainly appear in transferases, and phosphorylation is the most common type of modification found in allosteric proteins. Disorders related to allosteric proteins primarily comprise metabolic diseases and cancers. In general, allosteric proteins prefer to exist as monomers or even-numbered multimers. Greater stability and hydrophobicity are observed in allosteric proteins than in general proteins. Further analysis of the allosteric sites reveals a series of buried and compact pockets composed of significantly greater hydrophobic surface area than the corresponding orthosteric sites. The hydrophobicity of the allosteric sites plays a dominant role in the binding of allosteric modulators as observed in the analysis of 106 diverse allosteric protein–modulator pairs. These results may be of great significance in predicting which proteins are allosteric and in designing novel triggers to inhibit or activate proteins of interest.



## 1. Introduction

With the growing number and availability of genome sequences and gene expression profiles, more attention has been given to protein function and regulation in the post-genomic era [1]. As one of the most direct, rapid, and efficient methods for regulating protein function, allostery plays a key role in controlling biological activity, ranging from metabolic mechanisms to signal transduction pathways [2]. Allostery is defined as the regulation of protein function, structure, and/or flexibility induced by the binding of a ligand at a site topographically distinct from the orthosteric site; this site is then defined as an allosteric site. Allosteric behaviors are mainly induced by the specific binding of metal ions or molecules, resulting in the activation of cellular responses for homeostasis [3]. Anomalous regulation of allostery has been associated with a broad spectrum of human disorders and diseases including cancer, diabetes, Alzheimer's disease, and inflammation [4–6].

The earliest examples of allosteric regulation include the control of oxygen binding to hemoglobin by protons [7] and the AMP activation of glycogen phosphorylase [8]. The allosteric family has now expanded from multimeric proteins to monomeric proteins as well as from native proteins to engineered proteins [9–11]. In fact, most of allosteric proteins are accidentally discovered and the knowledge of allosteric mechanisms remains limited. Monod et al. first analyzed 24 allosteric enzyme systems and proposed a "plausible model on the nature of allosteric transition" ("MWC" model) in 1965 [12]. Koshland et al. challenged the MWC model and proposed the sequential hypothesis ("KNF" model) [13]. Both the MWC and KNF models are associated with multi-domain proteins and cannot be explained to recent observations of allostery in single domain proteins. Notably, the remarkable progress of X-ray crystallography and NMR methods in allosteric proteins has prompted the investigation of allosteric regulation from a structural point of view in recent years, and structures have been used as starting points for the study of protein dynamics, conformational equilibria, and

* Corresponding author at: Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai JiaoTong University, School of Medicine, Shanghai 200025, China.
Tel.: +86 21 63846590x776922; fax: +86 21 64154900.
*E-mail address:* jian.zhang@sjtu.edu.cn (J. Zhang).
[1] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.12.011

ultimately allosteric functions [14–16]. However, the structures of most allosteric proteins are still unavailable and as a result, the common features derived from these proteins are often underestimated or ignored [17]. Therefore, a systematic analysis of various allosteric proteins may provide a more thorough understanding of allosteric characteristics and functions.

Intrinsically, the allosteric regulation of a protein is triggered by the binding of a modulator to the allosteric site; the binding transmits the conformational change from the allosteric site to the orthosteric site via atom fluctuations, amino acid residue networking, or domain motion, eventually leading to a switching between two or more conformational states and, thus, functions [18–20]. In contrast to conventional orthosteric sites, allosteric sites offer a high level of selectivity, and modulators bound to them may mimic physiological conditions and result in fewer side effects when the chemotypes diverge from orthosteric ligands [21]. For example, the successes of Gleevec (an allosteric inhibitor of Abl) [22], Cinacalcet (an allosteric activator of calcium-sensing receptor) [23], and Maraviroc (an allosteric inhibitor of chemokine receptor 5) [23] promise exciting therapeutic prospects for new classes of drugs targeting allosteric sites. Some naturally occurring allosteric sites that affect the regulation of enzymes in central metabolic pathways have been extensively studied [24–26]. However, it is also of paramount importance to identify and characterize new allosteric sites [27]. As new allosteric sites are gradually exploited, interest in the development of empirical methods for predicting allosteric sites grows and the requirement for a statistical description of known allosteric sites is extremely urgent.

Although the importance of allosteric regulation has been realized, a general understanding of the structural basis of allosteric proteins and sites remains elusive due to the unsystematic organization of information about allostery. Recently, a major allosteric reference resource, the ASD database, has been built in our group to summarize a large number of allosteric proteins with explicit annotations of allosteric sites [17]. Using this collection, we report the results from a series of analyses of the sequence and structural characteristics of various allosteric proteins and their allosteric sites. Furthermore, preferences regarding the distribution of allosteric sites and modulator binding are also uncovered by statistical evaluation. The present study may provide a vital clue to understanding the ground rules for predicting allosteric sites and for designing novel allosteric proteins.

## 2. Materials and methods

### 2.1. Dataset

Two datasets of proteins were collected in the present study: an allosteric protein set and a common protein set. The raw source of the allosteric protein set was the AlloSteric Database (ASD, version 2010.8), a recent database that provides a central resource for allosteric molecules [17]. In the ASD, 512 allosteric proteins from 101 species were manually collected from the literature. To explore the characteristics of allosteric proteins, a set of ligand-binding proteins were built as a common control, which was taken from the Binding Database (BindingDB, version 2010) [28], a publicly accessible resource currently containing 6490 proteins with experimentally determined ligands. Because of the redundancy of the collection process of both original databases, a two-stage method was used to ensure the quality of the datasets in the study: (i) all identified allosteric proteins in the ASD were removed from the BindingDB dataset; (ii) proteins with high similarity (sequence identify $\geq 40\%$) were removed from each dataset using a Perl script [29]. Finally, 381 allosteric proteins in the ASD (hereinafter referred to as the allosteric set) and 3317 proteins in the BindingDB (hereinafter referred to as the common set) were retrieved. These proteins were used as samples for determining properties and for sequence analysis and were the source of subsets in the site and binding analysis.

To investigate the allosteric features in the allosteric proteins, two subsets of the allosteric set were built, viz., the interaction subset (Table S1) and the site subset (Table S2). In the allosteric set, 85 proteins complexed with their allosteric modulators were obtained from the Protein Data Bank (PDB). Considering that some allosteric proteins contain different allosteric modulators, only crystal structures with a low structural similarity (value < 0.5 in the FP2 method encoded in Openbabel, http://openbabel.org) were selected as subjects in the interaction subset, leading to 106 pairs from the 85 allosteric proteins. The site subset was constructed from the allosteric proteins in which both the orthosteric and allosteric pockets were known. The pockets were detected by the Ghecom algorithm (http://biunit.naist.jp/ghecom) [30] using the allosteric proteins in the interaction subset and validated by manual inspection according to the original literature. In the process, 52 allosteric proteins with the identified pockets were collected in the site subset for site analysis.

### 2.2. Phylogenetic tree

To outline the family profile of allosteric proteins, a phylogenetic tree was built from the allosteric set. Multiple sequence alignment (MSA) was carefully performed by the 'progressive' algorithm [31] among 381 proteins using ClustalX [32], and MEGA 5.0 [33] was used to build the phylogenetic tree from the MSA results.

### 2.3. Functional classification of allosteric proteins

To determine the common attributes of the allosteric proteins in biological processes, four functional classifications, viz., (i) primary enzyme commission (EC) classification; (ii) subcellular location; (iii) post-translational modifications (PTMs); (iv) associated disease, were described for the allosteric set. The primary EC numbers 1–6 (oxidoreductase, transferase, hydrolase, lyase, isomerase and ligase) for these proteins were annotated from UniProt [34] and Enzyme Nomenclature [35]. The PTMs and subcellular location of the proteins were taken from UniProt [34]. The diseases associated with the allosteric proteins were curated from OMIM [36] and then classified by the International Classification of Disease (ICD, http://apps.who.int/classifications/apps/icd/icd10-online/).

### 2.4. Sequence analysis

Six sequence descriptors widely used in the sequence analysis of proteins were calculated in both the allosteric and common sets: frequency of amino acid (FA), frequency of dipeptide (FD), frequency of amino acid group (FAG), hydrophobicity (HP), aliphatic index (AI) and instability index (II).

FA is the frequency of each amino acid in a protein, which was calculated by dividing the number of each of the 20 standard amino acids by the length of the protein. Similarly, FD is the frequency of dipeptides in a protein and was obtained by dividing the number of each of the 400 dipeptides ($20 \times 20$) by the length of each protein.

Amino acids in eukaryotes were grouped, as is common, according to their side-chains' structure and charge at a physiological pH of 7.4 as follows: (i) amino acids with electrically charged side chains (arginine, histidine, lysine, aspartic acid, and glutamic acid); (ii) amino acids with polar uncharged side chains (serine, threonine, asparagines, and glutamine); (iii) aromatic amino acids (phenylalanine, histidine, tryptophan, and tyrosine); (iv) amino acids with hydrophobic side chains (alanine, valine, isoleucine, leucine,

methionine, phenylalanine, tyrosine, and tryptophan); (v) special cases (cysteine, glycine, and proline). Then, the FAG was obtained by dividing the total number of amino acids in each group by the length of protein.

HP was used to characterize the global hydrophobicity of a protein. The hydrophobicity index for each amino acid was derived from the Kyte and Doolittle index [37], and HP was calculated by dividing the sum of the hydrophobicity indices of all of the amino acids in a protein by the length of protein.

AI is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine) in a protein and may be regarded as a positive factor for increasing the thermostability of globular proteins. The AI of a protein was calculated according to Eq. (1) [38]:

$$\mathrm{AI} = X(\mathrm{Ala}) + a \times X(\mathrm{Val}) + b \times (X(\mathrm{Ile}) + X(\mathrm{Leu})) \quad (1)$$

where $X$(Ala), $X$(Val), $X$(Ile) and $X$(Leu) are the mole percent (100 × mole fraction) of Ala (alanine), Val (valine), Ile (isoleucine), and Leu (leucine). The coefficients $a$ and $b$ are the volumes of Val side chain ($a$ = 2.9) and of Leu/Ile side chains ($b$ = 3.9), respectively, relative to the side chain of Ala.

II [39] can be used to predict a protein's instability and is based on the dipeptide instability weight value (DIWV) in Eq. (2):

$$II = \left(\frac{10}{L}\right) \sum_{i=1}^{L-1} \mathrm{DIWV}(x_i y_{i+1}) \quad (2)$$

where $x_i y_{i+1}$ is a dipeptide, $L$ is the length of the sequence and 10 is a scaling factor. A protein with an instability index smaller than 40 is considered to be a stable system, and a value above 40 indicates that the protein may be unstable.

### 2.5. Site analysis

For systematic understanding of the allosteric site, the features of several sites pertaining to composition, size, shape and flexibility as well as the distance between the allosteric and orthosteric sites in a protein were calculated using the allosteric proteins in the site subset.

The composition characteristics, such as FA, FAG and HP, of both the allosteric and orthosteric pockets were calculated with the same procedures used in the sequence analysis of proteins, using the residues in the pocket instead of the entire protein. The secondary structure around the pocket was captured using the 'Bioinfo' module in ICM (Molsoft LLC, San Diego, CA). The size and shape of each pocket were determined using the Ghecom method [30]. The distance between the allosteric and orthosteric pockets in a protein was calculated in two steps: (i) the center coordinate of a pocket was calculated as the arithmetic average of the coordinates of all the residues constituting the pocket; (ii) the distance between the centers of the allosteric and orthosteric pockets in the protein was measured.

The flexibility of the pocket was described using the temperature factors (B-factors) of the atoms that formed the residues in the pocket. To exclude variations between proteins, the B-factors were first taken from the PDB and then normalized to have a distribution with a mean of zero and unit variance [40,41] in each pocket using Eq. (3):

$$B' = \frac{B - \langle B \rangle}{\sigma(B)} \quad (3)$$

where $\langle B \rangle$ is the average value and $\sigma(B)$ is the standard deviation of all atoms in a protein. The flexibility of a pocket can be regarded as the average of B′ value of just the Cα atoms (normalized B-factor 1) and of all atoms (normalized B-factor 2) in the residues constituting the pocket.

### 2.6. Binding analysis

To determine the contributions of allosteric interactions between a protein and a modulator, a relevant description in terms of energy components should be taken into account. The hydrophobic effect, hydrogen bonding, and solvent accessible surface (SAS) percentage are required to describe their contributions to the allosteric binding.

X-SCORE [42] is an empirical scoring function that estimates the interaction between a protein and its ligand. X-SCORE has been used in several studies to reliably assess the contribution of energy items, including the hydrophobic effect and hydrogen bonding, to the protein–ligand interface [43–46]. Three individual scoring functions, the HSScore, HPScore, and HMScore, were implemented that included a van der Waals interaction term, a hydrogen bonding term, a hydrophobic effect term, a torsional entropy penalty, and a regression constant:

$$\mathrm{HSScore} = (C_{\mathrm{VDW},1})(\mathrm{VDW}) + (C_{\mathrm{H\text{-}bond},1})(\mathrm{HB}) + (C_{\mathrm{hydrophobic},1})(\mathrm{HS}) + (C_{\mathrm{rotor},1})(\mathrm{RT}) + C_{0,1}$$

$$\mathrm{HPScore} = (C_{\mathrm{VDW},2})(\mathrm{VDW}) + (C_{\mathrm{H\text{-}bond},2})(\mathrm{HB}) + (C_{\mathrm{hydrophobic},2})(\mathrm{HP}) + (C_{\mathrm{rotor},2})(\mathrm{RT}) + C_{0,2}$$

$$\mathrm{HMScore} = (C_{\mathrm{VDW},3})(\mathrm{VDW}) + (C_{\mathrm{H\text{-}bond},3})(\mathrm{HB}) + (C_{\mathrm{hydrophobic},3})(\mathrm{HM}) + (C_{\mathrm{rotor},3})(\mathrm{RT}) + C_{0,3}$$

The van der Waals term (VDW) was calculated using a softened Lennard–Jones 8–4 potential. The hydrogen binding term (HB) calculates all of the hydrogen bonds between ligand and protein with geometry-dependent functions. The rotor term (RT) calculates the number of "effective" rotors in the ligand molecule. These three terms are the same in all three scoring functions. For the hydrophobic effect term, HSScore calculates the buried hydrophobic molecular surface of the ligand (HS), HPScore calculates the pairwise hydrophobic atom contact potential (HP), and HMScore calculates the microscopic matching of hydrophobic ligand atoms to the binding pocket (HM). All three scoring functions were calibrated by reproducing the known binding affinities of 200 protein–ligand complexes. According to the procedure of X-SCORE, all crystal complexes in the interaction subset were separated with the proteins in the PDB format and the modulators in the MOL2 format. Water molecules and other cofactors in the proteins were removed from PDB files if they have no effect on hydrogen bonds between the protein and its modulator. Atom types and charges were assigned by default parameters in SYBYL-X 1.1 (Tripos Inc, St. Louis, MO, USA) for both proteins and modulators. Then, X-SCORE was used to calculate the hydrophobic effect (HMScore, HSScore and HPscore) and hydrogen bonding for each pair of allosteric proteins and modulators in the interaction subset.

SAS is the surface area accessible to a "water molecule". In the complex, SAS refers to the level of exposure of the ligand when bound to the site of the protein. Therefore, the SAS percentage, calculated by dividing the SAS of a ligand in the complex by the total surface of the ligand, could be an indicator of the affinity of the ligand and druggability of the site. Then, the SAS and the total surface of a ligand were determined using ICM (Molsoft LLC, San Diego, CA) on each protein–modulator pair in the interaction subset.

## 3. Results

The characteristics of the allosteric proteins were summarized in Table 1, and the features of the allosteric sites were listed in Table 2.

### 3.1. Phylogenetic tree

In the last decade, major progress has been made in the discovery of allosteric proteins. To profile the allosteric proteins, a phylogenetic tree (Fig. S1) was constructed using the proteins in the allosteric set. The tree shows a wide phylogenetic distribution of these allosteric proteins, indicating the adequate diversity of our allosteric set. Furthermore, both the interaction subset and the site subset were labeled (as red and purple dots, respectively) in the phylogenetic tree, which indicates that these sets were representative of the allosteric proteins used for site analysis and binding analysis.

### 3.2. Functional classification of allosteric proteins

The functions of allosteric protein are crucial to the control of many biological processes in various subcellular locations. In this study, we investigated the functions of allosteric proteins by classifying them according to four criteria: enzyme class, subcellular location, post-translational modification (PTMs) and disease. As described in Fig. 1, almost half of the known allosteric enzymes are transferases (44.8%), and the next most common types of allosteric enzymes are oxidoreductases (20.9%) and hydrolases (20.3%). The subcellular distribution shows that allosteric proteins are most often found in the membrane (40.0%) and cytoplasm (25.2%). Meanwhile, PTMs on the known allosteric proteins were analyzed; the results indicate that the predominant type of post-translational modification in allosteric proteins is phosphorylation (34.8%). In addition, glycosylation (28.8%) and acetylation (21.3%) are also observed in a large number of allosteric proteins. Finally, dysfunctions of most allosteric proteins are related to diseases that occurred in the metabolic (18.6%) and nervous (12.8%) systems or were cancerous (13.6%), and nervous system.

### 3.3. Sequence analysis

To explore the size of allosteric proteins, the distributions of protein length (the number of amino acids) were calculated for both the allosteric and common sets (Fig. 2A). The result shows that no obvious difference in size was found between the allosteric proteins and common proteins. Approximately 80% of allosteric proteins are composed of 200–800 amino acids.

Allostery was frequently found in proteins with multiple subunits, and the preferred number of subunits could be viewed as a composition feature of allosteric proteins. The distribution of this composition property in the allosteric set was shown in Fig. 2B. Aside from the monomers, proteins with even numbers of subunits are observed significantly more often than those with an odd number of subunits, indicating the importance of pairing in the composition of the allosteric proteins containing multiple subunits.

FA (frequency of amino acid) is a common sequence descriptor to depict the physicochemical composition of a protein. The frequencies of 20 amino acids were calculated in Fig. 2C for both the allosteric and common sets. According to the group classification described in the FAG descriptor (see Section 2 for details), the FA preference in the allosteric proteins could be described in terms of five characteristics: (i) four of the five electrically charged amino acids show higher (aspartic acid, $P$-value < 0.01; glutamic acid, $P$-value < 0.05; lysine, $P$-value < 0.01) or lower (histidine, $P$-value < 0.01) occupancies in the allosteric set; (ii) three of the four polar uncharged amino acids display larger (asparagine, $P$-value < 0.01) or smaller (serine, $P$-value < 0.01; glutamine, $P$-value < 0.01) distributions in the allosteric proteins; (iii) three of the four aromatic amino acids (phenylalanine, histidine and tryptophan) exhibit lower distributions in the allosteric set at the level of $P$-value < 0.01; (iv) six of the eight hydrophobic amino acids show larger (alanine, $P$-value < 0.05; valine, $P$-value < 0.01; isoleucine, $P$-value < 0.01) or smaller (leucine, $P$-value < 0.01; phenylalanine, $P$-value < 0.01; tryptophan, $P$-value < 0.01) occupancies in the allosteric proteins; (v) both cysteine and proline show significantly lower distributions in the allosteric set at the level of $P$-value < 0.01. Interestingly, of all the amino acids, the largest composition differences between the allosteric and common datasets were found for the hydrophobic isoleucine and the polar serine. Isoleucine shows a higher occupancy in the allosteric set, whereas the frequency of serine is larger in the common set.

In comparison with the FA, which focuses on a single type of amino acid, the FAG (frequency of amino acid group) can provide a global view for evaluating the nature of a protein by physiochemical groups. Five groups were compared between the allosteric and the common sets. As tabulated in Table 1, the allosteric proteins have more electrically charged groups and fewer aromatic and special groups than the common proteins ($P$-value < 0.01).

The FD (frequency of dipeptide) serves as a complementary sequence descriptor for FA and FAG and can help to sharply delineate the relationships among amino acids in a given protein. In this analysis, the frequencies of 400 dipeptides were calculated for both the allosteric set and the common set. Next, to avoid the deficiency of some dipeptides in proteins, 121 dipeptides with FD values at the top half of both datasets were compared, leading to 47 significantly different dipeptides with $P$-value < 0.01. Among the 47 dipeptides shown in Fig. 2D, the first 37 (Group I) have larger occupancies in the allosteric proteins, and the remaining 10 (Group II) show smaller occupancies. Furthermore, 86.5% of the dipeptides in Group I have no less than one hydrophobic amino acid, whereas only 60.0% of the dipeptides in Group II have hydrophobic amino acids, indicating the significance of hydrophobic amino acids for allosteric protein. Moreover, those high-frequency hydrophobic dipeptides

**Table 1**
The comparative properties of proteins in the allosteric and control sets.

| Properties | Allosteric proteins | | Control proteins | | $P$-value[a] |
|---|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation | |
| Electrically charged group (%) | 25.14 | 4.56 | 24.38 | 4.86 | <0.01 |
| Polar uncharged group (%) | 20.09 | 3.23 | 20.38 | 3.47 | 0.120 |
| Aromatic group (%) | 10.83 | 2.64 | 11.64 | 2.80 | <0.01 |
| Hydrophobic group (%) | 41.32 | 4.29 | 40.84 | 5.29 | 0.085 |
| Special group (%) | 13.44 | 3.21 | 14.40 | 3.70 | <0.01 |
| Hydrophobicity | −0.18 | 0.29 | −0.22 | 0.34 | 0.035 |
| Aliphatic index | 89.41 | 11.81 | 86.38 | 13.54 | <0.01 |
| Instability index | 39.13 | 9.15 | 41.83 | 9.95 | <0.01 |

[a] The $P$-values were calculated using Origin 8.0. The $P$-values less than 0.01 are all shown as <0.01.

**Table 2**
The comparative properties of pockets in the site subset.

| Properties | Allosteric pockets | | Orthosteric pockets | | *P*-value[a] |
|---|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation | |
| Electrically charged group (%) | 24.01 | 9.256 | 25.05 | 6.271 | 0.501 |
| Polar uncharged group (%) | 18.23 | 8.19 | 20.97 | 5.368 | 0.045 |
| Aromatic group (%) | 11.96 | 5.677 | 12.77 | 4.768 | 0.429 |
| Hydrophobic group (%) | 44.02 | 10.72 | 38.36 | 8.189 | 0.032 |
| Special group (%) | 13.65 | 6.588 | 15.58 | 5.000 | 0.093 |
| Hydrophobicity | −0.016 | 0.661 | −0.349 | 0.450 | <0.01 |
| Normalized B-factor 1 | −0.065 | 0.535 | −0.077 | 0.396 | 0.906 |
| Normalized B-factor 2 | −0.066 | 0.495 | −0.070 | 0.364 | 0.967 |
| Pocket size | 1145.72 | 1314.29 | 3265.84 | 2959.87 | <0.01 |
| Pocket depth | 4.66 | 1.138 | 4.65 | 1.034 | 0.866 |

[a] The *P*-values were calculated using Origin 8.0. The *P*-values less than 0.01 are all shown as <0.01.

may contribute to build either hydrophobic allosteric pocket or motion hinge in the allosteric proteins. Remarkably, isoleucine is the most frequent amino acid (48.6%) in Group I, and serine (20%) shows the highest frequency in Group II; these findings are in good agreement with the previous analysis of the FA.

At the molecular level, hydrophobicity plays a vital role in driving protein folding, insertion of membrane proteins into the nonpolar lipid environment, and enhancing protein–small molecule interactions. As described above, several hydrophobic amino acids and dipeptides have a significant impact on the composition of the allosteric proteins. Analyzing the hydrophobic group in the FAG, however, does not show an evident difference between the allosteric set and the common set (*P*-value = 0.085 in Table 1). Considering the oversimplified categories in the FAG, a more accurate descriptor, HP, was introduced to evaluate the overall hydrophobicity of proteins between the two datasets (see Section 2 for details). The result in Table 1 reveals that the allosteric proteins have higher HP values than the common proteins with *P*-value < 0.05, indicating that the allosteric proteins are more hydrophobic. This finding is in good agreement with what has recently uncovered by Namboodiri et al. [47] that hydrophobic patterning plays a pivotal role in determining allostery in terms of comparing the allosteic protein set derived from our ASD database [17] with a randomly generated sequence populations as well as a genetic protein set.

The AI (aliphatic index) is a positive predictor for the thermostability of globular proteins. Table 1 shows that the allosteric proteins have a higher AI than the common proteins (89.41 versus 86.38)

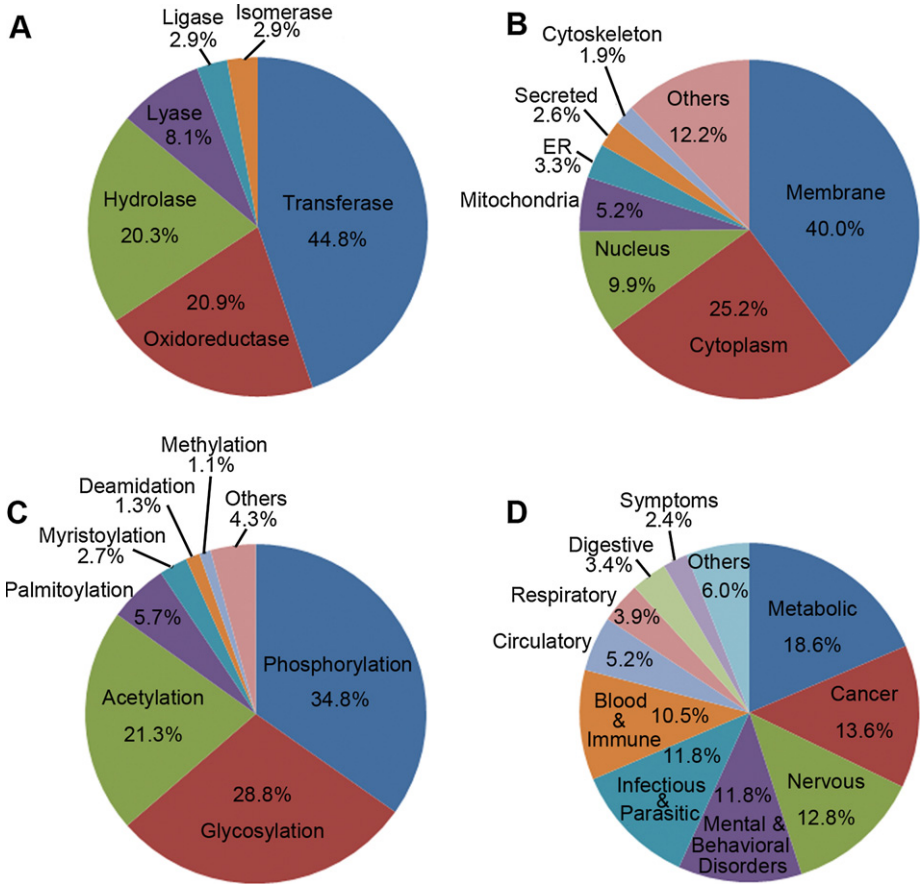


**Fig. 1.** Functional classification of allosteric proteins. (A) Primary enzyme commission (EC) classification. (B) Subcellular location classification. (C) Post-translational modifications (PTMs) classification. (D) Associated disease classification.

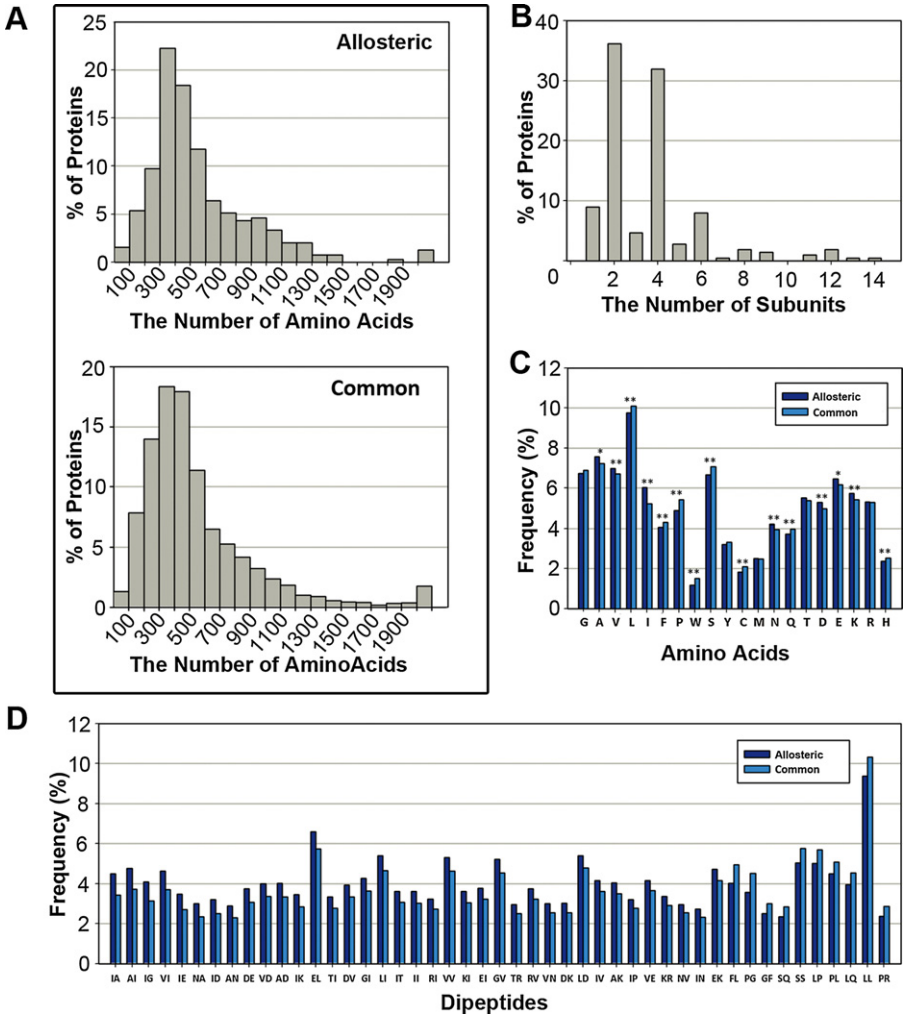


**Fig. 2.** Sequence analysis of allosteric proteins. (A) Distributions of protein length in both the allosteric and common sets. (B) Distribution of preferred number of subunits in allosteric proteins. (C) Frequency of amino acid (FA) in both the allosteric and common sets. * and ** indicate that the property between the allosteric and common sets is significantly different with $P$-value < 0.05 and $P$-value < 0.01, respectively. (D) Frequency of dipeptide (FD) in both the allosteric and common sets. All types of dipeptides between the two sets are different with $P$-value < 0.01. These two-sample $t$-tests were calculated using Origin 8.0.

with $P$-value < 0.01. This finding may suggest that the allosteric proteins tend to have greater thermostability and could be better at survival and transfer in a variety of cellular micro-environments.

The II (instability index) reflects the instability of a protein and might be used as another important descriptor in the analysis of protein stability. Compared with the higher II values found in the common proteins, the mean II value of less than 40 in the allosteric proteins provides statistical evidence ($P$-value < 0.01) that these proteins have greater stability and a longer half-life (Table 1). Moreover, the II finding is also consistent with the result of AI.

### 3.4. Site analysis

Allosteric regulation occurs when a modulator binds to the allosteric site, which is topographically distinct from the substrate pocket. Therefore, it could be beneficial to examine the allosteric sites involved in the allosteric regulation as well as their composition and properties on a wider scale. In addition, the relationship between the allosteric site and the orthosteric site should be determined.

First, the FA and FAG of orthosteric sites and allosteric sites from the proteins in site subset were calculated (Fig. 3A and Table 2). More isoleucine residues were found in the allosteric sites than in the orthosteric sites ($P$-value < 0.01). Moreover, the allosteric sites have a higher proportion of hydrophobic groups (44.02%) than the orthosteric sites ($P$-value < 0.01). This finding shows the dominant contribution of the hydrophobic groups and is consistent with the hydrophobic difference calculated using HP values between the two types of sites ($P$-value < 0.01, Table 2). Nevertheless, only 18.23% of the amino acids that have polar groups were detected in the allosteric sites, and this value is less than that in the orthosteric sites ($P$-value < 0.05). These results suggest that hydrophobicity, rather than polar interactions, could play a more important role in allosteric binding than in substrate binding. In addition, glycine is found more frequently in the orthosteric sites ($P$-value < 0.05), and the decreased number of glycine residues in the allosteric sites may restrict some movements that frequently take place in the orthosteric sites.

The B-factors in the crystal structures were used as a measure of residue flexibility [40,48]. In the present study, mean normalized B-factors (B') were calculated for the two types of sites (Table 2). The residues of the allosteric sites tend to have slightly higher B-factors than those in the orthosteric sites for when considering just the Cα atoms (normalized B-factor 1) or all atoms (normalized B-factor 2). However, no significant difference was observed in the analysis, indicating that both types of sites may have similar flexibility.

The binding sites are mostly cavities or grooves composed of several elements of secondary structures, which are often used to

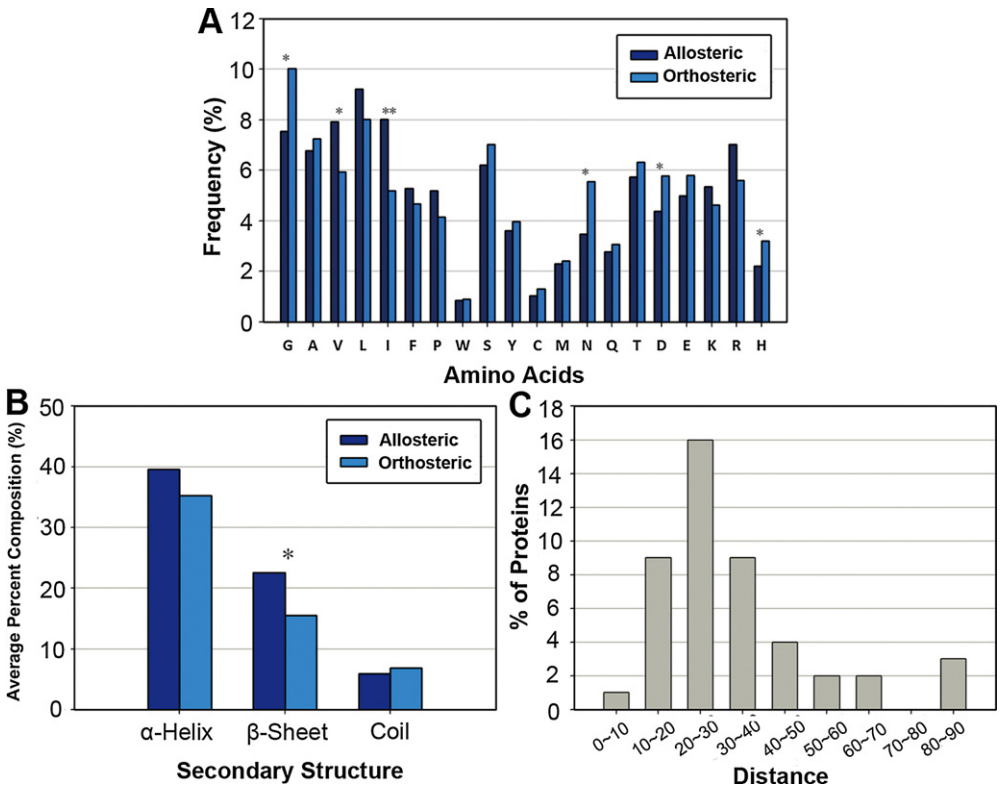


**Fig. 3.** Site analysis of allosteric sites and their corresponding orthosteric sites. (A) Frequency of amino acid (FA) in both allosteric sites and orthosteric sites. (B) Comparison of the distributions of three types of secondary structures in both allosteric sites and orthosteric sites. (C) Distribution of the distances between allosteric sites and their corresponding orthosteric sites. * and ** indicate that the property between the allosteric sites and the orthosteric sites is significantly different with $P < 0.05$ and $P < 0.01$, respectively. These two-sample *t*-tests were calculated using Origin 8.0.

determine the relationship between structure and function [49]. Fig. 3B shows the secondary structure assignment of residues in the allosteric sites and the orthosteric sites. As expected, α-helices and β-sheets are frequently found in both types of sites with a higher proportion (39.60% compared to 22.58%) in the allosteric sites. At the same time, fewer random coils occur in the allosteric sites, reflecting the more ordered construction of allosteric sites. Thus, the more ordered structure in the allosteric sites could provide a specific conformational transition and facilitate the allosteric communication from allosteric site to orthosteric site.

The size and shape of a site is one of crucial factors in deciding which type of ligand to bind. Table 2 lists the pocket volume and depth of the allosteric sites and the orthosteric sites of proteins from the site subset. The allosteric sites tend to have smaller volumes than the orthosteric sites (1145.72 Å$^3$ versus 3265.84 Å$^3$, *P*-value < 0.01) whereas there is no obvious difference between the average depths of the two groups of sites. This result could be due to the different number of ligands bound in a site; the catalytic function of orthosteric sites in enzymes such as transferases generally occurs via the binding of two substrates and even additional cofactors, which require more space for enclosure. Based on the literature, 78.9% of the proteins in the site subset are indeed able to bind more than one substrate in the orthosteric sites, but all allosteric proteins have a single modulator in the allosteric sites.

Allosteric behavior acts as a type of signal transduction between two discrete patches within a macromolecule, and analyzing the distance from the start to terminus may allow a more systematic and accurate interpretation of various allosteric mechanisms. Fig. 3C shows the distribution of distances from the allosteric sites to the orthosteric sites in allosteric proteins from the site subset. The shortest distance is 6.69 Å and observed in the *Escherichia coli* L-asparaginase I. The longest path of 87.35 Å is found in the *Methanococcus jannaschii* threonine-sensitive aspartokinase. Within this range, a normal distribution for the distances is proposed, and more than 90% of the distances between the allosteric sites and the orthosteric sites are from 10 to 50 Å. Because of this distribution, the emergence of an allosteric site might only depend on its local structural character and reliable communication to the orthosteric site instead of a specified distance from the orthosteric site.

### 3.5. Binding analysis

The analysis indicates that there are distinctive properties of allosteric proteins and their allosteric sites. Therefore, understanding how these properties relate to the binding of modulators could be of great value in the discovery of allosteric drugs. Using a well-known empirical scoring function, X-SCORE, the contribution of various energy items to the allosteric binding was evaluated with 106 allosteric protein–ligand pairs in the interaction subset. Among the energy items, hydrogen bonding and the hydrophobic effect are two major contributors, and the average percent composition of hydrogen bonding and hydrophobic effect was calculated as shown in Fig. 4A. The result shows that the hydrophobic effect dominates the allosteric interaction to a greater extent than hydrogen bonding, especially in non-endogenous allosteric protein–ligand pairs (Fig. 4B). This finding is in good agreement with our finding of more hydrophobic and fewer polar patches in the site analysis, as well as consistent with our previous study of the molecular basis for allosteric modulators [50]. Taken together, these observations definitely corroborate that binding is predominantly determined by the presence of hydrophobic patches in the allosteric sites.

Desolvation is a major factor driving a ligand into its receptor [51–55], can be evaluated by the SAS (solvent accessible surface) percentage of a ligand in the binding site, and reflects the ability of the pocket to shield small molecules from the solvent and

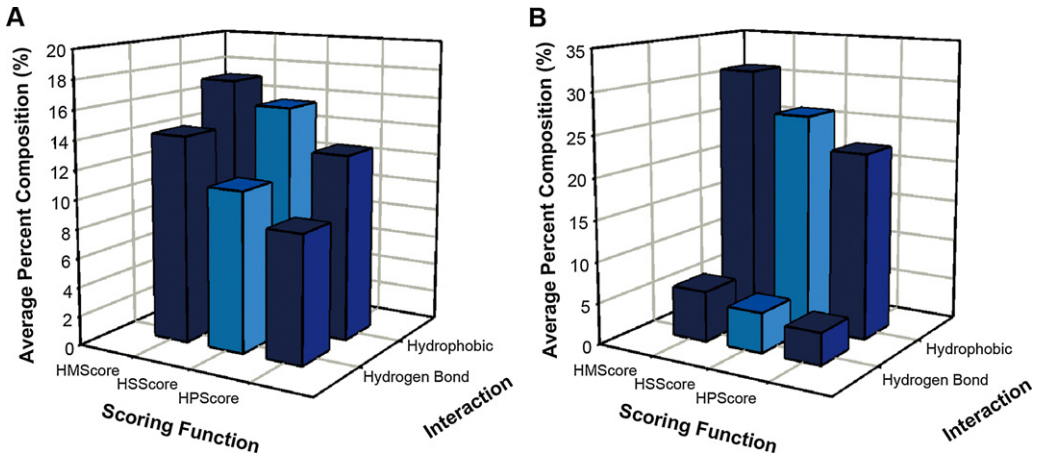


**Fig. 4.** Comparison of the contribution of hydrophobic effect and hydrogen bonding in allosteric proteins using X-SCORE. (A) The interaction subset containing 106 protein–modulator pairs. (B) The 44 exogenous protein–modulator pairs found by removing the endogenous interactions from the interaction subset.

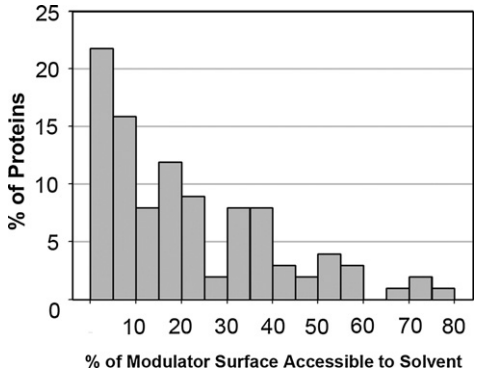


**Fig. 5.** Distribution of SAS percentage in allosteric proteins from the interaction subset.

hydrophobicity/hydrophilicity of the ligand [56–59]. A low SAS percentage indicates that a ligand has less surface area accessible to the "water molecule" and more areas that are buried in the site, which has been frequently found in the cases of highly hydrophobic ligands. As shown in Fig. 5, a skewed distribution was found when analyzing the SAS percentage and most allosteric modulators have low SAS percentages. These findings suggest that allosteric sites can bury small molecules and that allosteric modulators have a hydrophobic potential.

## 4. Discussion

Four functional classifications of the allosteric proteins were exhibited in Fig. 1; Fig. 1A shows that transferases represent 44.8% of all allosteric proteins. This value is distinct from the distribution in another type of ligand-binding protein–drug targets. Previous studies revealed that drug targets are more likely to be oxidoreductases (33%), followed by transferases (31%) [60]. This difference indicates that ligands could have a greater impact on regulating transferases by allosteric mechanisms. In terms of subcellular location, most drug targets are on the membrane (57%), and some are in the cytoplasm (13%) [60]. In contrast, fewer allosteric proteins are found in the membrane (40.0%), and more allosteric proteins tend to locate in the cytoplasm (25.2%), which may be attributed to the spatial requirement for large-scale dynamic movements in some allosteric proteins (Fig. 1B). The fact that phosphorylation constitutes the largest proportion (34.8%) of PTMs in allosteric proteins indicates its close association with allostery (Fig. 1C). Interestingly, the dependence of phosphorylation on allostery has been observed in the *Drosophila* homeodomain proteins Even-skipped [61] and StyR [62]. Most endogenous allosteric ligands were found during the peak period of metabolic research [1], and it is not surprising that metabolic diseases are most common in the allosteric-related disorders (Fig. 1D).

Most multimeric allosteric proteins analyzed in the study have an even number of subunits. Using biological annotation of the allosteric proteins, we found that some of their functions are based on having a symmetric structure (e.g., acetyl-CoA carboxylase and epidermal growth factor receptor) [63,64] and some are based on having a pair of complementary units—one regulatory subunit and one catalytic subunit (e.g., 5′-AMP-activated protein kinase and isocitrate dehydrogenase) [65,66]. This finding could partially explain this composition preference in multimeric allosteric proteins. Numerous sequence properties indicate a series of specific features of the allosteric proteins as shown in Table 2. Combining these results with the FA, FD, FAG and HP analysis indicates that the allosteric proteins show a more hydrophobic trend than the common proteins as a whole. Particularly, isoleucine and serine may become residue markers to differentiate allosteric proteins from common proteins. From the AI and II results, one could deduce that allosteric proteins exhibit a more stable character than common proteins, which could allow a long lifetime for maintaining dynamic regulation in the cellular micro-environment.

Allosteric sites are where signal transduction originates in allosteric proteins, and signal transduction is triggered by the binding of allosteric modulators. However, it is not a trivial matter to find new allosteric sites and glean novel allosteric modulators for the sites due to a lack of empirical research. Our site analysis and binding analysis have investigated the fundamental principles for allosteric binding. In general, the volumes of allosteric pockets are nowhere near as large as those of substrate pockets, and all known allosteric pockets can only accommodate one molecule. In contrast, the mean depth and flexibility of allosteric pockets appear to be nonspecific, but allosteric modulators in the pockets have ratios that favor burial based on the SAS percentage values. These observations depict a concave and compact space for the allosteric sites, providing a global view for detecting allosteric surfaces. Moreover, the physicochemical properties inside the allosteric sites illustrate the local functional components for driving allosteric binding. More hydrophobic and less polar distributions in the composition of the allosteric sites lead to prominent contributions of hydrophobic interactions in the allosteric binding. This binding is clearly different from the substrate binding in orthosteric sites

and has been demonstrated in both statistical analysis and case studies. Similar findings have also been previously observed in the studies of caspase-7, GlyP, and β-lactamase by Hardy and Wells [27]. Furthermore, De Ruvo et al. [67] developed the graph theory method to differentiate the allosteric and non-allosteic process. The results show that the peculiar feature of allosteric communication, to be effective, is a change of active site residues contacts while non-allosteric regulation is not. Taken together, by systematically exploring a large number of examples above, the trends in both the landscape of allosteric sites and the modes of allosteric binding could become clearer and more detailed, providing the ground rules for allosteric identification. These ground rules may serve as the foundation for future design strategies targeting the development of allosteric systems.

In the future, using the summary of allosteric properties from the study, a knowledge-based method could be introduced to develop a novel algorithm for the prediction of allosteric sites, providing valuable opportunities for drug discovery. Furthermore, the annotation of all potential allosteric proteins throughout the human genome would help to reveal the entire landscape of allostery.

## Acknowledgments

We would like to thank Prof. Hanyi Zhuang in Shanghai Jiaotong University for polishing the manuscript. This work was supported in part by grants from the National Basic Research Program of China (973 Program) (2012CB967001), the National Natural Science Foundation of China (21002062, 21102090), a Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201083), Shanghai Jiao Tong University K. C. Wong Medical Fellowship Fund, the Innovative Research Team of Shanghai Municipal Education Commission, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, the Shanghai PuJiang Program (10PJ406800), and the China Postdoctoral Science Foundation (2012M520046).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.12.011.

## References

[1] J.E. Lindsley, J. Rutter, Whence cometh the allosterome? Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 10533–10535.
[2] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nature Chemical Biology 4 (2008) 474–482.
[3] D. Eisenberg, E.M. Marcotte, I. Xenarios, T.O. Yeates, Protein function in the post-genomic era, Nature (2000) 823–826.
[4] A. Taly, P.J. Corringer, D. Guedin, P. Lestage, J.P. Changeux, Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system, Nature Reviews Drug Discovery 8 (2009) 733–750.
[5] G. Munro, P.K. Ahring, N.R. Mirza, Developing analgesics by enhancing spinal inhibition after injury: GABAA receptor subtypes as novel targets, Trends in Pharmacological Sciences 30 (2009) 453–459.
[6] F.M. Matschinsky, Assessing the potential of glucokinase activators in diabetes therapy, Nature Reviews Drug Discovery 8 (2009) 399–416.
[7] Proceedings of The Physiological Society: January 22, 1910. The Journal of Physiology 40 (1910) i–vii.
[8] G.T. Cori, S.P. Colowick, C.F. Cori, The action of nucleotides in the disruptive phosphorylation of glycogen, Journal of Biological Chemistry 123 (1938) 381–389.
[9] K. Kamata, M. Mitsuya, T. Nishimura, J. Eiki, Y. Nagata, Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase, Structure 12 (2004) 429–438.
[10] J. Lee, M. Natarajan, V.C. Nashine, M. Socolich, T. Vo, W.P. Russ, S.J. Benkovic, R. Ranganathan, Surface sites for engineering allosteric control in proteins, Science 322 (2008) 438.
[11] A. Villaverde, Allosteric enzymes as biosensors for molecular diagnosis, FEBS Letters 554 (2003) 169–172.
[12] J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model, Journal of Molecular Biology 12 (1965) 88–118.
[13] D. Koshland Jr., G., N. methy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1966) 365–385.
[14] D.D. Boehr, D. McElheny, H.J. Dyson, P.E. Wright, The dynamic energy landscape of dihydrofolate reductase catalysis, Science 313 (2006) 1638–1642.
[15] V.J. Hilser, E.B. Thompson, Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins, Proceedings of the National Academy of Sciences of the United States of America 104 (2007) 8311.
[16] J. Zhang, C. Li, K. Chen, W. Zhu, X. Shen, H. Jiang, Conformational transition pathway in the allosteric process of human glucokinase, Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 13368–13373.
[17] Z. Huang, L. Zhu, Y. Cao, G. Wu, X. Liu, Y. Chen, Q. Wang, T. Shi, Y. Zhao, Y. Wang, ASD: a comprehensive database of allosteric proteins and modulators, Nucleic Acids Research 39 (2011) D663–D669.
[18] M.D. Daily, J.J. Gray, Local motions in a benchmark of allosteric proteins, Proteins 67 (2007) 385–399.
[19] M.D. Daily, T.J. Upadhyaya, J.J. Gray, Contact rearrangements form coupled networks from local motions in allosteric proteins, Proteins 71 (2008) 455–466.
[20] B.A. Kidd, D. Baker, W.E. Thomas, Computation of conformational coupling in allosteric proteins, PLoS Computational Biology 5 (2009) e1000484.
[21] J.A. Lewis, E.P. Lebois, C.W. Lindsley, Allosteric modulation of kinases and GPCRs: design principles and structural diversity, Current Opinion in Chemical Biology 12 (2008) 269–280.
[22] M.A. Bogoyevitch, D.P. Fairlie, A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding, Drug Discovery Today 12 (2007) 622–633.
[23] P.J. Conn, A. Christopoulos, C.W. Lindsley, Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, Nature Reviews Drug Discovery 8 (2009) 41–54.
[24] J. Grimsby, R. Sarabu, W.L. Corbett, N.E. Haynes, F.T. Bizzarro, J.W. Coffey, K.R. Guertin, D.W. Hilliard, R.F. Kester, P.E. Mahaney, Allosteric activators of glucokinase: potential role in diabetes therapy, Science 301 (2003) 370.
[25] V.L. Rath, M. Ammirati, D.E. Danley, J.L. Ekstrom, E.M. Gibbs, T.R. Hynes, A.M. Mathiowetz, R.K. McPherson, T.V. Olson, J.L. Treadway, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chemistry and Biology 7 (2000) 677–682.
[26] C. Wiesmann, K.J. Barr, J. Kung, J. Zhu, D.A. Erlanson, W. Shen, B.J. Fahr, M. Zhong, L. Taylor, M. Randal, Allosteric inhibition of protein tyrosine phosphatase 1B, Nature Structural & Molecular Biology 11 (2004) 730–737.
[27] J.A. Hardy, J.A. Wells, Searching for new allosteric sites in enzymes, Current Opinion in Structural Biology 14 (2004) 706–715.
[28] X. Chen, Y. Lin, M. Liu, M.K. Gilson, The binding database: data management and interface design, Bioinformatics 18 (2002) 130–139.
[29] L. Holm, C. Sander, Removing near-neighbour redundancy from large protein sequence collections, Bioinformatics 14 (1998) 423–429.
[30] T. Kawabata, Detection of multiscale pockets on protein surfaces using mathematical morphology, Proteins 78 (2010) 1195–1211.
[31] D.F. Feng, R.F. Doolittle, Progressive sequence alignment as a prerequisite to correct phylogenetic trees, Journal of Molecular Evolution 25 (1987) 351–360.
[32] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Research 25 (1997) 4876–4882.
[33] S. Kumar, M. Nei, J. Dudley, K. Tamura, MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences, Briefings in Bioinformatics 9 (2008) 299–306.
[34] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, The universal protein resource (UniProt), Nucleic Acids Research 33 (2005) D154–D159.
[35] IUBMB, Enzyme Nomenclature, Acad. Press, San Diego, CA, 1992.
[36] A. Hamosh, A.F. Scott, J.S. Amberger, C.A. Bocchini, V.A. McKusick, Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders, Nucleic Acids Research 33 (2005) D514–D517.
[37] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, Journal of Molecular Biology 157 (1982) 105–132.
[38] I. Atsushi, Thermostability and aliphatic index of globular proteins, Biochemical Journal 88 (1980) 1895–1898.
[39] K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Engineering 4 (1990) 155–161.
[40] Z. Yuan, J. Zhao, Z.X. Wang, Flexibility analysis of enzyme active sites by crystallographic temperature factors, Protein Engineering 16 (2003) 109–114.
[41] O. Carugo, P. Argos, Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors, Proteins 31 (1998) 201–213.
[42] R. Wang, L. Lai, S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction, Journal of Computer-Aided Molecular Design 16 (2002) 11–26.
[43] T. Cheng, X. Li, Y. Li, Z. Liu, R. Wang, Comparative assessment of scoring functions on a diverse test set, Journal of Chemical Information and Modeling 49 (2009) 1079–1093.

[44] R. Wang, Y. Lu, X. Fang, S. Wang, An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes, Journal of Chemical Information and Computer Science 44 (2004) 2114–2125.
[45] R. Wang, Y. Lu, S. Wang, Comparative evaluation of 11 scoring functions for molecular docking, Journal of Medicinal Chemistry 46 (2003) 2287–2303.
[46] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, Protein Engineering 8 (1995) 127–134.
[47] S. Namboodiri, A. Giuliani, A.S. Nair, P.K. Dhar, Looking for a sequence based allosteric definition: a statistical journey at different resolution scales, Journal of Theoretical Biology 304 (2012) 211–218.
[48] S.-Y. Lu, Y.-J. Jiang, J.-W. Zou, T.-X. Wu, Molecular modeling and molecular dynamics simulation studies of the GSK3β/ATP/substrate complex: understanding the unique P+4 primed phosphorylation specificity for GSK3β substrates, Journal of Chemical Information and Modeling 51 (2011) 1025–1036.
[49] R. Samudrala, Structural bioinformatics, New Jersey: Wiley-Liss, Briefings in Bioinformatics 4 (2003) 299–301.
[50] Q. Wang, M. Zheng, Z. Huang, X. Liu, H. Zhou, Y. Chen, T. Shi, J. Zhang, Toward understanding the molecular basis for chemical allosteric modulator design, Journal of Molecular Graphics and Modelling 38 (2012) 324–333.
[51] S.-Y. Lu, Y.-J. Jiang, J. Lv, J.-W. Zou, T.-X. Wu, Mechanism of kinase inactivation and nonbinding of FRATide to GSK3β due to K85M mutation: molecular dynamics simulation and normal mode analysis, Biopolymers 95 (2011) 669–681.
[52] J. Janin, C. Chothia, Role of hydrophobicity in the binding of coenzymes, Biochemistry 17 (1978) 2943–2948.
[53] S.-Y. Lu, Y.-J. Jiang, J.-W. Zou, T.-X. Wu, Effect of double mutations K214/A-E215/Q of FRATide on GSK3β: insights from molecular dynamics simulation and normal mode analysis, Amino Acids 43 (2012) 267–277.
[54] S.-Y. Lu, Y.-J. Jiang, J. Lv, J.-W. Zou, T.-X. Wu, Role of bridging water molecules in GSK3β-inhibitor complexes: insights from QM/MM MD, and molecular docking studies, Journal of Computational Chemistry 32 (2011) 1907–1918.
[55] S. Lu, W. Huang, X. Li, Z. Huang, X. Liu, Y. Chem, T. Shi, J. Zhang, Insights into the role of magnesium triad in *myo*-inositol monophosphatase: metal mechanism substrate binding, and lithium therapy, Journal of Chemical Information and Modeling 52 (2012) 2398–2409.
[56] T.A. Halgren, Identifying and characterizing binding sites and assessing druggability, Journal of Chemical Information and Modeling 49 (2009) 377–389.
[57] S. Henrich, O.M.H. Salo-Ahen, B. Huang, F.F. Rippmann, G. Cruciani, R.C. Wade, Computational approaches to identifying and characterizing protein binding sites for ligand design, Journal of Molecular Recognition 23 (2010) 209–219.
[58] R.P. Sheridan, V.N. Maiorov, M.K. Holloway, W.D. Cornell, Y.D. Gao, Drug-like density: a method of quantifying the bindability of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank, Journal of Chemical Information and Modeling 50 (2010) 2029–2040.
[59] S.-Y. Lu, Y.-J. Jiang, J.-W. Zou, T.-X. Wu, Molecular modeling and molecular dynamics simulation studies on pyrrolopyrimidine-based α-helix mimetic as dual inhibitors of MDM2 and MDMX, Journal of Molecular Graphics and Modelling 30 (2011) 167–178.
[60] T.M. Bakheet, A.J. Doig, Properties and identification of human protein drug targets, Bioinformatics 25 (2009) 451–457.
[61] C. Li, J.L. Manley, Allosteric regulation of even-skipped repression activity by phosphorylation, Molecular Cell 3 (1999) 77–86.
[62] M. Milani, L. Leoni, G. Rampioni, E. Zennaro, P. Ascenzi, M. Bolognesi, An active-like structure in the unphosphorylated StyR response regulator suggests a phosphorylation-dependent allosteric activation mechanism, Structure 13 (2005) 1289–1297.
[63] Y. Shen, S.L. Volrath, S.C. Weatherly, T.D. Elich, L. Tong, A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product, Molecular Cell 16 (2004) 881–891.
[64] X. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kuriyan, An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor, Cell 125 (2006) 1137–1149.
[65] B. Xiao, R. Heath, P. Saiu, F.C. Leiper, P. Leone, C. Jing, P.A. Walker, L. Haire, J.F. Eccleston, C.T. Davis, Structural basis for AMP binding to mammalian AMP-activated protein kinase, Nature 449 (2007) 496–500.
[66] A.B. Taylor, G. Hu, P.J. Hart, L. McAlister-Henn, Allosteric motions in structures of yeast NAD+-specific isocitrate dehydrogenase, Journal of Biological Chemistry 283 (2008) 10872–10880.
[67] M. De Ruvo, A. Giuliani, P. Paci, D. Santoni, L. Di Paola, Shedding light on protein–ligand binding by graph theory: the topological nature of allostery, Biophysical Chemistry 165-166 (2012) 21–29.