Biophysics and Physicobiology

Regular Article

Dynamic profile analysis to characterize dynamics-driven allosteric sites in enzymes

Junko Taguchi[1] and Akio Kitao[2]

[1]*Tsukuba Research Center, Taiho Pharmaceutical Co., Ltd, Tsukuba, Ibaraki, Japan*
[2]*Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo 113-0032, Japan*

We examine the dynamic features of non-trivial allosteric binding sites to elucidate potential drug binding sites. These allosteric sites were previously found to be allosteric after determination of the protein-drug co-crystal structure. After comprehensive search in the Protein Data Bank, we identify 10 complex structures with allosteric ligands whose structures are very similar to their functional forms. Then, possible pockets on the protein surface are searched as potential ligand binding sites. To mimic ligand binding to the pocket, complex models are generated to fill out each pocket with pseudo ligand blocks consisting of spheres. Normal mode analysis of the elastic network model is performed for the complex models and unbound structures to assess the change of protein dynamics induced by ligand binding. We examine nine profiles to describe the dynamic and positional characteristics of the pockets, and identify the change of fluctuation around the ligand, ΔMSF_{bs}, as the best profile for distinguishing the allosteric sites from the other sites in 8 structures. These cases should be considered as examples of dynamics-driven allostery, which accompanies significant changes in protein dynamics. ΔMSF_{bs} is suggested to be used for the search of potential dynamics-driven allosteric sites in proteins for drug discovery.

Corresponding author: Akio Kitao, Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan.
e-mail: kitao@iam.u-tokyo.ac.jp

Key words: protein, allosteric ligand, elastic network model, normal mode

Allosteric drug binding sites have drawn recent attention as new types of drug targets. From the early 2000s, a particular type of ligand binding site has been experimentally discovered in various proteins through drug discovery research [1]. Ligands of this type of binding site had not been recognized as a new type prior to identification of the site. These ligands were found to bind the sites distant from functionally important endogenous ligand binding sites, while typical drugs directly block such ligand binding sites. This new type of allosteric drug binding site can be considered as an attractive target to obtain new intellectual property and/or to overcome drug resistance. Only a few allosteric drug binding sites of this type have been discovered to date, because they were typically found by chance using a set of expensive experiments, e.g., the combination of high throughput screening of a large compound library and determination of the complex structure by X-ray crystallography. The successes of such experiments are not necessarily guaranteed, because they are a matter of chance. Therefore, a theoretical prediction method is required to compensate for experimental difficulties.

To predict the possible ligand binding sites of proteins,

◄ *Significance* ►

The dynamic features of non-trivial allosteric binding sites were examined to elucidate potential drug binding sites. After comprehensive search in the Protein Data Bank, we identified 10 complex structures with allosteric ligands that do not cause significant conformational change. 101 complex models were generated to fill out possible pockets on the protein surface and normal mode analysis of the elastic network model was performed to examine the change of protein dynamics induced by ligand binding. We found the change of fluctuation around the ligand as the best profile for distinguishing the allosteric sites from the others.
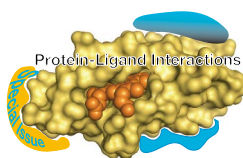
many different profiles, e.g., shape of the pockets, amino acid composition, and binding free energy, have been examined. Ming *et al.* proposed a dynamics perturbation analysis (DPA) algorithm to predict functional sites in protein structures [2–4]. The regions where interactions cause a large change in the protein conformational distribution were found using relative entropy. They succeeded in predicting 267 binding sites out of 305 proteins [2] from the GOLD [5] docking test set [6]. They also analyzed a functionally allosteric protein, trypsinogen, using a method similar to DPA and concluded that relatively strong communication between the regulatory and active sites was evident [4]. Demerdash *et al.* used a support vector machine to distinguish allosteric sites of functionally allosteric enzymes, transcription factors, and signal transduction proteins [7].

Allostery comes from the Greek *allos+steric* (other+ space), which indicates the regulation of reactions by ligand binding to a site distant from the active site. Allostery has been studied for a long time and was found to widely regulate enzymatic reactions and signal transductions. To understand the complexity of allostery, the following four aspects are typically considered. 1) Protein structure: classic examples of allostery have been found in multimeric proteins, such as in the well-known Monod-Wyman-Changeux (MWC) [8] or Koshland-Némethy-Filmer (KNF) [9] models, whereas allostery is now recognized as a property observed in monomeric proteins as well as in the multimeric protein. 2) Conformational change: Tsai *et al.* classified allosteric protein structures that do not involve a change of backbone shape [10]. According to their analysis, some proteins do not significantly change their conformation with the binding of allosteric ligands. In some kinases, well-known allosteric sites are formed with conformational change of the loop near the active site. 3) Effect: Two opposite methods of regulation, positive or negative, depending on whether the effecter increases or decreases the ligand-binding affinity of the protein. 4) Regulator: The ligands can be either homotropic or heterotropic, and either small compounds or proteins. Homotropic is where the allosteric ligand (modulator or effecter) is identical to the active site ligand, whereas heterotopic is where the ligands are different. Not only small compounds, but also proteins are known as allosteric ligands. In this respect, protein-protein interaction as well as protein-small compound interaction should be included in allosteric regulation.

In 2004, Gunasekaran *et al.* stated that allostery is not limited to well-known oligomeric proteins, but is more likely to be found as an intrinsic property of all proteins, and that structural perturbation at any site could lead to redistribution of the conformational sub-states [11]. This concept is based on the idea that allosteric ligands do not necessarily cause large structural changes, which suggests tight coupling between allosteric ligand binding and the change induced in the protein dynamics. Redistribution of a protein conformational ensemble, referred to as dynamics-driven allostery, is a powerful concept to rationalize allosteric regulation.

In this work, we examine the dynamic characteristics of the dynamics-driven allosteric sites in proteins and attempt to quantify features common among the allosteric sites by analyzing the fluctuation of the simulated proteins. Ligand-complex co-crystal structures were selected from the Protein Data Bank [12] and possible binding pockets on the protein surface were searched. Pseudo ligand blocks (PLBs), which mimic binding ligands, were generated to fill the pockets. Elastic network model (ENM) calculations [13] of the complex models constructed using a protein monomer and a PLB were performed to examine the effects of the ligands on the protein dynamics. Eight profiles that are expected to describe the characteristics of the allosteric regulation were selected and examined. After analyzing the difference among the allosteric, active, and other possible ligand binding sites, quantities that better discriminated the experimentally identified allosteric sites were identified. The dynamics-driven allosteric sites can be considered to be structurally sensitive to perturbation, which induces significant changes in the protein dynamics upon ligand binding.

Material and Methods

Selection of dynamics-driven allosteric proteins

In this work, we focused on monomeric proteins without significant conformational change upon ligand binding. Proteins that satisfied the following three conditions were selected: 1) the functional unit is monomeric, 2) the allosteric site was not known to be functionally important before the complex structure was solved, and 3) the regulator is a drug-like organic compound. In most of the multimeric allosteric proteins, large conformational changes occur among domains and ligand binding to the allosteric sites located at the domain-domain or protein-protein interface interferes with functionally important movement. Thus, we assumed that there could be differences in mechanisms between monomer allosteric proteins and multimers.

To define the dataset for the analysis, protein structures were searched on the RSCB Protein Data Bank (PDB) website [12]. The following three conditions were applied to searching queries: 1) Has Ligands, 2) keyword: alloste (-ric, -rically, -ry), and 3) Resolution: ≤ 2.5 angstrom. 890 allosteric co-crystal structures that satisfied the three criteria were selected from the PDB website: "Has Ligands", "keyword: allosteric" and "Resolution: ≤ 2.5 angstrom". The PDB was retrieved on May 8th, 2012. 4326 ligands contained in these co-crystal structures were also downloaded in SMILES form. The obtained structures were further screened using three additional conditions applied to all including ligands with Pipeline pilot [14]: 1) organic compounds, 2) drug-likeness to meet the Lipinski's Rule of five [15], and 3) Molecular weight: >200. Among the 4398 ligand structures constructed, 3376 ligands survived the applied first two conditions. The third condition was applied to eliminate

Table 1 Selected proteins co-crystallized with allosteric ligands

Type	Protein[a]	UNP	Ligand Activity	# of complex models	RMSD[a] from apo structure (Å)	RMSD[a] from active site complex (Å)
Enzyme	GlmU [29] (2VD4)	P43889	18 μM (IC50)	5	0.21 (2V0H)	0.15 (2V0J)
	glucokinase [32] (1V4S)	P35557	Activate 15 fold at 1 μM	8	9.2 (1V4T)	0.74 (3IDH)
	HCV NS5B.1 [33] (2HWI)	P26663	3 μM (IC50)	18	0.48 (1C2P)	0.37 (3BSC)
	HCV NS5B.2 [30] (2BRK)	P26663	26 nM (IC50)	20	1.1 (1C2P)	1.1 (3BSC)
	CK2 [34] (3H30)	P68400	40 μM (Ki)	8	1.1 (1NA7)	1.2 (1JWH)
	p38 MAP kinase.1 [17] (1KV1)	Q16539	5900 nM (IC50)	11	1.2 (1WFC)	1.1 (1ZYJ)
	p38 MAP kinase.2 [16] (3HVC)	Q16539	600 nM (IC50)	10	1.4 (1WFC)	0.77 (1ZYJ)
	TEM1 [35] (1PZP)	P62593	490 μM (Ki)	7	0.94 (1YT4)	0.93 (1AXB)
	PTP1B [36] (1T48)	P18031	350 μM (IC50)	7	0.86 (3SME)	0.96 (2CM7)
Receptor	Androgen receptor [37] (2PIP)	P10275	Low-affinity	7	no available structure	0.24 (2AMA)

[a] PDB ID is indicated in parenthesis.

typical non-specific ligands. Finally 736 co-crystal structures were obtained. These were classified into 179 UniProt families and subsequently eliminated as "functionally allosteric protein", "oligomeric protein" and those structures without reference papers, which resulted in the selection of 9 co-crystal structures. One additional structure of p38 MAP kinase, p38 MAP kinase.2 hereafter (PDB ID: 3HVC [16]), was also found when p38 MAP kinase.1 (PDB ID: 1KV1 [17]) was investigated. Although the ligand is not explicitly described as allosteric, this was determined to be the case after examining the structure and the original paper [16]. In each case, it was confirmed that the ligand was bound to a site different from the active site by comparing the PDB entries with identical UniProt IDs. Thus, a total of 10 structures were selected for this work (Table 1); Nine enzyme and one receptor protein structures comprised of 8 distinct proteins. The co-crystal structures with known active site ligands were also collected. Two new allosteric sites were found for HCV NS5B (HCV NS5B.1 and HCV NS5B.2, hereafter) and p38 MAP kinase (the aforementioned p38 MAP kinase.1 and p38 MAP kinase.2). Only one receptor protein in the dataset is an androgen receptor that contains hormone-like compounds in the cofactor site. Large collective conformational change upon ligand binding (hinge bending) was found only in glucokinase (see root-mean-square deviations, RMSD, in Table 1). Glucokinase is also the only protein where the ligand functions as an activator. In the cases of HCV NS5B.2 and p38 MAP kinase.1, partial conformational change is supposed to induce the formation of a new allosteric binding pocket upon ligand binding be-

cause these allosteric sites do not exist in either the active site co-crystal or the apo structures. Of note, the allosteric site of p38 MAP kinase.1 overlaps with the active site, because the allosteric sites were formed by the conformational change around the active site. In the other cases, the allosteric sites are distant from the active sites.

The structures of the selected proteins bound to the allosteric ligands are very similar to those of the active site complexes, i.e., RMSD are around 1 Å or smaller (Table 1). This indicates that the allosteric effects on these proteins are not evident from the structure change only but are rather expected to be related to the change in protein dynamics (Fig. 1A, B). Therefore, we examined the effects of the ligand binding on protein dynamics in this work.

Elastic network model

The effects of the ligand binding were analyzed by the elastic network model (ENM) and normal mode analysis. ENM is a coarse-grained model for proteins [13] in which all the interactions between pairs of heavy atoms within a cutoff distance are approximated by the harmonic potential and the protein dynamics are described by normal modes as a linear combination of harmonic oscillators.

In this work, the ENM was employed to analyze protein fluctuations for the following two reasons. Firstly, the ENM is suitable to observe the dynamic effects of a protein with a simple and fast calculation [13]. In ENM, collective normal modes are determined only by diagonalization of a Hessian matrix of the potential energy. A more accurate molecular dynamics (MD) approach is much more time consuming,
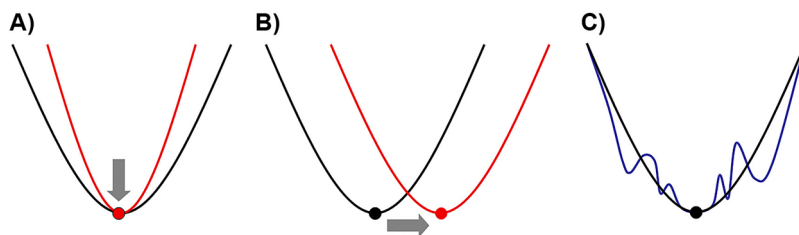
Figure 1 Basic concept of the dynamics-driven allostery and ENM visualized by effective energy surface as a function of conformational change. A) Energy surface change in dynamics-driven allostery which accompanies change in conformational distribution. Possible potential energy surfaces are unbounded (black line) and bound (red line) structures. The stable structure does not change upon ligand binding. B) The change in classic allostery. Conformational change occurs upon ligand binding, which corresponds to the shifts of the stable structure. C) Potential energy surface in the fine-grained all-atom model (blue line) and coarse-grained ENM (black line). The middle dot indicates a stable structure (typically a crystal structure).

because the equation of motion is to be solved step by step. It is well known that the collective motions of proteins are well described with the ENM, despite its rough approximation of the interaction energy and shorter calculation time [18,19]. The ENM is expected to provide a good approximation to smooth out the fine structures of the all-atom potential energy surface (Fig. 1C). Secondly, all heavy atoms are treated equally under the coarse-graining approximation; carbon, nitrogen, oxygen, and any other heavy atoms are not distinguished and the effects of the ligand on the protein dynamics are examined regardless of the detailed chemical structure of the ligand.

Construction of complex models

Complex structure models based on ENM were constructed using the following procedure. Unliganded model structures were first prepared by removing all the heteroatoms (ligands and water molecules) from the original co-crystal structures of the protein-allosteric ligand complexes. Pockets on the surface of the unliganded models were searched using the PASS [20] program to find possible ligand binding sites. The binding sites found were then filled with probe spheres. Active site points (ASPs), regarded as the centers of binding pockets, were obtained. The blocks of the probes were generated to fill out the pocket by gathering the probes within a distance of 6 Å around each ASP. The radius of the PASS probe is smaller than the radius of carbon and the density of the probes are much higher than typical chemical ligands; therefore, pseudo ligand blocks (PLBs) to fill out the pockets were re-generated as follows. Firstly, the volume of the first blocks was calculated using the MAMA program [21], and the number of applicable atoms to fill the volume was then calculated. An appropriate number density value of 0.09 Å^{-3} was employed from analysis of the protein and ligands of this dataset. Finally, the coordinate of PLBs were generated using the Situs program [22]. Each combination of a PLB and unliganded protein was treated as a protein-ligand complex model and was employed for the following analysis. The complex models with PLB in the allosteric and active sites are termed allosteric site and active

site models, respectively. The other complex models are referred to as decoy models (see Fig. 2 for examples).

Normal mode analysis and calculation of MSF

To examine the effects of ligand binding to protein dynamics, ENM analysis was applied to the generated complex models and unliganded models. Tirion's all heavy atom model [13] was used. Under the harmonic approximation, the potential energy is defined as:

$$E = \sum_{d_{ij}^0 < R_c} C(d_{ij} - d_{ij}^0)^2 , \qquad (1)$$

where d_{ij} is the distance between atom i and j, d_{ij}^0 is the distance between the corresponding atoms in the reference structure, R_c is the cutoff distance, and C is the strength of the potential (force constant). Here, the reference structure indicates each model structure, R_c is set to 9 Å and C is set to 1 (unitless). Therefore, the magnitude of the atomic fluctuation is significant only when the difference or ratio is considered. After diagonalization of the Hessian matrix $\mathbf{H}$, normal mode eigenvectors and eigenvalues are obtained,

$$\mathbf{U^T H U} = \mathbf{\Lambda} , \qquad (2)$$

where $\mathbf{U}$ is the eigenvector matrix whose k^{th} column vector $\mathbf{u}_k$ is the kth eigenvector. The kth element of Λ, λ_k is equal to ω_k^2, where ω_k is the angular frequency of the kth normal mode. The set of vectors is orthonormal and linearly independent. $\mathbf{u}_{ki}$ is a three-dimensional subset of $\mathbf{u}_k$ for the ith atom.

The mean square fluctuation (MSF) of each atom is obtained from the eigenvalues and eigenvectors of the ENM. MSF of the ith atom with the kth mode is given by,

$$MSF_{ki} = \frac{k_B T \mathbf{u}_{ki}^T \mathbf{u}_{ki}}{\lambda_k} , \qquad (3)$$

where k_B is the Boltzmann constant and T is the absolute temperature. Thus MSF for the ith atom is obtained,

$$MSF_i = \sum_{k=1}^n MSF_{ki} . \qquad (4)$$
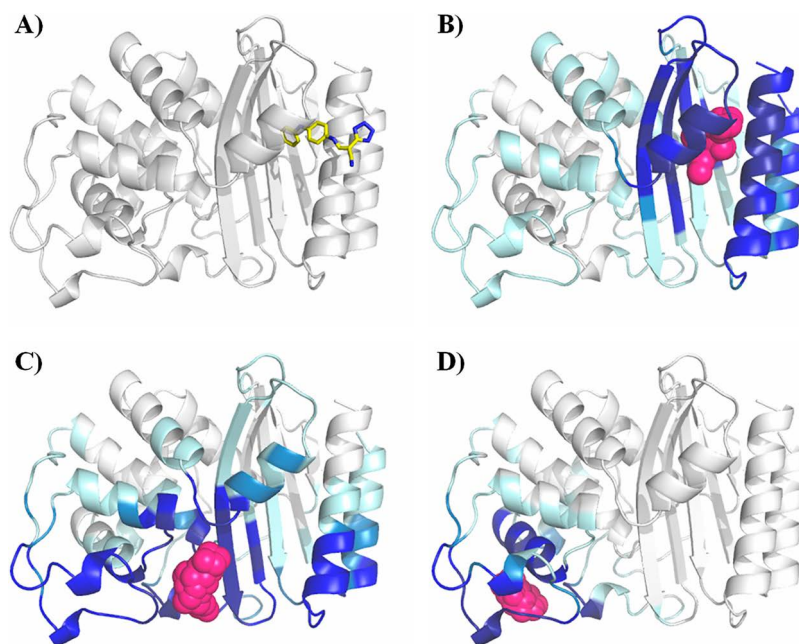
Figure 2 Examples of generated complex models. A) The complex structure of TEM1 (PDB ID: 1PZP), B) allosteric site model, C) active site model, and D) one of the decoy models. PLBs are shown by magenta spheres. Amino acid residues are colored according to their ΔMSF_i upon PLB binding; strong reduction in fluctuation (deep blue), medium reduction (cyan), and no significant reduction (white).

The summation in Eq. (4) was taken for the lowest 100 normal modes except for the zero frequency modes for overall translations and rotations. MSF of each residue was calculated as the average over the heavy atoms. The dynamic influence of each PLB per residue was assessed by subtracting MSF of the unliganded model from MSF of the corresponding complex model with PLB:

$$\Delta MSF_i = MSF_i^{\text{PLB}} - MSF_i^{\text{unliganded}}. \tag{5}$$

ζ_k is the reduction ratio of MSF along the kth normal mode upon the ligand binding, which is defined similar to the anharmonicity factor to quantify the fluctuation ratio between the principal and normal modes along each principal mode [23,24]. When ζ_k is unity, the fluctuation of the kth normal mode does not change. The condition $\zeta_k < 1$ indicates a decrease of the mode fluctuation and $\zeta_k > 1$ indicates the opposite. The total reduction ratio is defined as,

$$\zeta = \sum_{k=1}^{n} \zeta_k. \tag{6}$$

The lowest 100 normal modes were considered for this calculation.

Analyzed profiles

We analyzed the nine profiles shown in Table 2 as possible quantities correlated to allostery. The radius of gyration of PLBs (R_g) was selected to characterize the size of the ligands, which was calculated using the MAMA [21] program from Uppsala Software Factory. With R_g, we can examine if the size of the pocket filled with PLB correlates with the allosteric effect. Mean-square fluctuation of the PLB binding sites before binding (MSF_{bs}) was measured to examine if each binding site is intrinsically flexible in the

Table 2 Nine profiles examined to characterize dynamic allosteric sites

Profile	Definition
R_g	Radius of gyration of PLB
MSF_{bs}	MSF of the PLB binding site before binding
r_{cm-PLB}	Distance between the center of mass of protein and PLB
r_{as-PLB}	Distance between the active site and PLB
ΔMSF_{as}	ΔMSF around active sites
ΔMSF_{bs}	ΔMSF around the PLB binding site
ΔMSF_{all}	ΔMSF of all the residues
ζ	Reduction ratio of MSF
Druggability	Pocket druggability probability predicted by PockDrug-Server [25]

Table 3   Average values of the normalized nine profiles for the allosteric site, active site, decoy, and all models

Profile	Allosteric site[a]	Active site[b]	Decoy[b]	All[b]
R_g	0.62±0.31	0.63±0.31	0.44±0.31	0.49±0.32
MSF_{bs}	**0.42±0.36**	**0.59±0.37**	**0.63±0.30**	**0.60±0.31**
r_{cm-PLB}	0.59±0.32	0.32±0.36	0.58±0.31	0.55±0.33
r_{as-PLB}[c]	0.49±0.30	–	0.51±0.33	0.50±0.32
ΔMSF_{as}[c]	0.81±0.30	–	0.78±0.33	0.79±0.33
ΔMSF_{bs}	**0.42±0.36**	**0.59±0.37**	**0.63±0.30**	**0.60±0.31**
ΔMSF_{all}	0.50±0.31	0.40±0.41	0.63±0.30	0.59±0.32
ζ	**0.46±0.34**	**0.61±0.42**	**0.66±0.29**	**0.63±0.31**
Druggability	0.90±0.11	0.64±0.36	0.45±0.35	0.52±0.37

[a] Mean values of the 10 allosteric site models are shown. [b] Differences of the mean values from the allosteric site models are shown. [c] The active site models are excluded from the normalization.
Values shown before and after ± represent the average and standard deviation of the corresponding data, respectively.

unliganded state. MSF_{bs} was calculated for the residues within 4 Å from PLBs. To quantify the position of PLBs on the protein, the distance between the center of mass of the protein and PLB, $r_{pro-PLB}$ was introduced. We also analyzed the distance between the active site and PLB, r_{as-PLB}. With this quantity, we checked if the allosteric effect is related to the distance from the active site. The center of the active site was defined as the center of mass around the C_α atoms of the active site residues. ΔMSF_{as} was introduced to examine the change in atomic fluctuations around the active site upon ligand binding. The change of the average MSF of the residues within 4.0 Å from the active site ligands was considered between the unliganded and liganded proteins. r_{as-PLB} and ΔMSF_{as} were analyzed to investigate the influence of the ligand binding on the active sites. ΔMSF_{bs} quantifies the middle-range effect of the ligand binding on MSF. The average MSF of the residues within 30 Å from PLB was considered to get the difference. This value was found to most effectively discriminate the allosteric sites from the other sites among the examined nine profiles (see Results). With shorter cut off distances, we observed weaker correlations with the allostery. We also employed a profile to cover whole protein. ΔMSF_{all} shows the long-range effect of the ligand binding on MSF of all the residues. ζ defined by Eq. (6) quantifies the reduction ratio of MSF. Finally we examined druggability of the generated models by PockDrug-Server, which can extract the binding pocket from a submitted protein-ligand model and predict the pocket druggability as probability [25].

Results

For the 10 selected structures, 101 complex models were constructed (Table 2). The breakdown is: 10 active site models, 11 allosteric site models, one cofactor site model (TEM1) and 79 decoy site models. Two allosteric site models were generated for glucokinase. Root-mean-square deviations (RMSDs) of C_α atoms were calculated using MOE 2010.10 [26].

Table 4   Normalized MSF_{bs} for the allosteric site models

Type	Protein	Allosteric site models	
		MSF_{bs}	Rank/all
Enzyme	GlmU	0.00	1/5
	glucokinase[a]	0.59, 1.00	5, 8/8
	HCV NS5B.1	0.22	8/18
	HCV NS5B.2	0.61	6/20
	CK2	0.66	7/8
	p38 MAP kinase.1	0.00	1/11
	p38 MAP kinase.2	0.46	6/10
	TEM1	0.09	3/7
	PTP1B	0.73	6/7
Receptor	Androgen receptor	0.06	2/7

[a] glucokinase has two allosteric site models.

By definition the druggability ranges from zero to one. The raw values of the other eight profiles for ten protein structures were in the ranges, R_g: 1.4~3.7 Å, MSF_{bs}: 0.06~0.25, r_{cm-PLB}: 4.24 ~ 28.69 Å^2, r_{as-PLB}: 1.07~46.85 Å, ΔMSF_{as}: −0.5788~0.0007, ΔMSF_{bs}: −0.0092~−0.0004, ΔMSF_{all}: −3.822~−0.093, and ζ: 90.77~99.98 (Supplementary Table S1). The eight profiles were linearly normalized between 0 and 1 from the minimum to maximum values for each protein. Average values of the profiles are shown in Table 3 (see Supplementary Table S2 for the individual values with normalization). Large difference between the allosteric site models from the others is considered as good indicators to distinguish. Therefore, MSF_{bs}, ΔMSF_{bs}, ζ, and *druggability* are regarded to be relatively good indicators, which are analyzed more in detail below.

MSF_{bs} of the allosteric site models is shown in Table 4. As shown in Table 3, the average value of MSF_{bs} was smaller than that of the other sites, which means that the binding sites of the allosteric ligands are relatively less flexible regions compared to the other binding pockets on average. In GlmU, HCV NS5B.1, p38 MAP kinase.1, TEM1, and Androgen receptor, MSF_{bs} is significantly smaller than the other sites. However, this is not the feature common to all the

Table 5 Normalized ΔMSF_{bs} for the allosteric site models and original co-crystal complexes with the allosteric ligands

| Type | Protein | Allosteric site models | | | | Original |
		ΔMSF_{bs}	Rank/all[a]	*Druggability*	*Overlap ratio*	ΔMSF_{bs}[b]
Enzyme	GlmU	1.00	5/5	0.98	0.24	0.64(2)
	glucokinase	0.00, 0.09	1,2/8	0.91, 0.98	0.39, 0.87	0.29
	HCV NS5B.1	0.33	3/18	0.69	0.58	0.23
	HCV NS5B.2	0.61	4/20	1.00	0.42	0.23(2)
	CK2	0.26	2(1)/8	0.93	0.35	−0.05
	p38 MAP kinase.1	0.70	9/11	1.00	0.62	0.79
	p38 MAP kinase.2	0.32	3(2)/10	0.89	1.00	0.60
	TEM1	0.32	2/7	1.00	0.39	0.22
	PTP1B	0.00	1/7	0.87	0.79	−0.41
Receptor	Androgen receptor	1.00	7/7	0.70	0.67	1.00

[a] The number in parenthesis is the rank order excluding the active site model. [b] The number in parenthesis is the ranking within the original ligand complex and pseudo complex models.

cases. Interestingly, the catalytic sites of 98 non redundant enzymes are situated in dynamic minima simulated by Gaussian network model [27] and principal component analysis of solution structures of enzymes also showed the catalytic sites are highly immobile [28]. The present result implies that some of the allosteric sites have a feature similar to the catalytic sites.

ΔMSF_{bs} values of individual cases are shown in Table 5. For the enzymes, six cases in Table 5, the allosteric site models were found to be within the top 3. Even the two cases in which ΔMSF_{bs} of the allosteric site models were not significantly different from the others (GlmU and HCV NS5B.2), ΔMSF_{bs} calculated for the original complex structures with the allosteric ligands were ranked as the top 2. This difference was probably due to the fact that the PLBs were too small compared to the actual allosteric ligands, which also suggests the importance of the ligand size in allostery (see Discussion). Including the results from the original complex structure with the allosteric ligands, ΔMSF_{bs} works very well (8 out of 9 enzymes), except for the case of p38 MAP kinase.1, in which the allosteric site occupies a part of the active site. Therefore, p38 MAP kinase.1 should be understood as the exceptional case. ΔMSF_{bs} is not a good measure in the case of androgen receptor. Average ΔMSF_{bs} values for the 8 allosteric site models except for p38 MAP kinase.1 and PTP1B are 0.33±0.30, which is significantly smaller than the average value of 0.65±0.29 for 66 decoy blocks, indicating a significant reduction of protein fluctuation around the allosteric site caused by ligand binding. Therefore, ΔMSF_{bs} can be considered as the best indicator to characterize the allosteric site. As mentioned in Materials and Methods, the medium range distance of 30 Å was used for the calculation of ΔMSF_{bs}. As shown by ΔMSF_{all}, the long range effect showed weaker correlation with the allosteric effect. We also examined the range shorter than 30 Å but the present definition was found to be the best value to obtain the highest correlation with the allosteric sites.

In Table 5, the druggability is also shown. As shown in

Table 3, the druggability of the allosteric and active site models are relatively high, which is consistent to the fact that the ligands are stably bound to these sites. Another important feature is that the druggability of the allosteric site models are notably higher than the others (around 0.7 or higher) including the active site models, which indicates the druggability is an important factor to be considered to predict potential dynamics-driven allosteric sites. Overlap ratio shown in Table 5 is considered in Discussion.

ζ is a quantity that is expected to be a good indicator from the average value shown in Table 3. Table 6 shows individual ζ values. In the cases of GlmU, glucokinase, CK2, and TEM1, ζ values were smaller than 0.30 and the ranks of the allosteric site models were within the top 3; however, this index cannot clearly discriminate the allosteric site models from the other 6 cases. In addition, ζ values for the original complex structure with the allosteric ligands were not particularly good.

Before conducting the profile analysis, we had expected that the change of fluctuation around the active sites upon binding to the allosteric site would be significant; however, the average ΔMSF_{as} value for the allosteric sites shown in Table 3 was comparable to those of the other sites. Table 7 shows the details of ΔMSF_{as}. ΔMSF_{as} is significantly small only in the case of p38 MAP kinase.1, intermediate in GlmU and PTP1B, and is greater than 0.9 in all the other cases. This suggests that the dynamic change around the active sites is not strongly related to the allosteric effect in most of these cases. A clear exception is p38 MAP kinase.1, the only case where ΔMSF_{bs} is not a good indicator to distinguish the allosteric site among all the enzymes examined. As mentioned above, this is the case in which the allosteric site occupies a part of the active site, which cannot be considered as a typical allosteric site. The allosteric mechanism of p38 MAP kinase.1 is expected to be related to a clear reduction of MSF around the active site (ΔMSF_{as} for the original allosteric ligand is −0.44), which is suggested to be different from the mechanisms for the other sites.

Table 6　Normalized ζ for the allosteric site models and original co-crystal complexes with the allosteric ligands

Type	Protein	Allosteric site models		Original
		ζ	Rank/all [a]	ζ
Enzyme	GlmU	0.00	1/5	0.53
	glucokinase	0.00, 0.44	1, 4(3)/8	0.38
	HCV NS5B.1	0.64	5/18	0.53
	HCV NS5B.2	0.64	7/20	0.64
	CK2	0.07	3(2)/8	−0.26
	p38 MAP kinase.1	0.91	8/11	0.92
	p38 MAP kinase.2	0.67	4/10	0.74
	TEM1	0.28	3/7	0.37
	PTP1B	0.53	3(2)/7	0.38
Receptor	Androgen receptor	0.91	5/7	0.95

[a] The number in parenthesis is the rank order excluding the active site model.

Table 7　Normalized ΔMSF_{as} for the allosteric site models and original co-crystal complexes with the allosteric ligands

| Type	Protein	Allosteric site models		Original
		ΔMSF_{as}	Rank/All [a]	ΔMSF_{as}
Enzyme	GlmU	0.64	2/4	0.26
	glucokinase	1.00, 0.99	6,5/7	0.99
	HCV NS5B.1	1.00	14/17	1.00
	HCV NS5B.2	0.99	13/19	0.98
	CK2	0.93	3/7	0.95
	p38 MAP kinase.1	0.00	1/10	−0.44
	p38 MAP kinase.2	1.00	6/9	0.99
	TEM1	0.91	3/6	0.99
	PTP1B	0.50	2/6	0.44
Receptor	Androgen receptor	0.97	4/6	0.97

[a] The rank order excludes the active site models.

Discussion

The results of 8 cases out of the 10 examined structures in this work clearly show that these allosteric effects are correlated to reduction of the protein fluctuation around the allosteric site. Therefore, these cases can be considered as examples of dynamics-driven allostery. In a DPA study [4], the authors analyzed the allosteric mechanisms of trypsinogen and found relatively strong communication between the regulatory and active sites. In the proteins studied in this work, communication between the allosteric and active sites was not found to be necessarily significant except for p38 MAP kinase.1 because ΔMSF_{as} of the allosteric sites are comparable to the other sites. Similar to the DPA study with the GOLD test set [2], the ligands binding on the allosteric sites examined in this work are suggested to cause a change in the protein conformational distribution, as shown in Figure 1A, rather than causing an effect specific to the active sites. The values of ζ were found to be significant in some cases. Ligand binding to these allosteric sites causes a large change of the low frequency normal mode fluctuation, which indicates the effects are not localized around the allosteric sites but are distributed over the protein including the binding sites and allosteric sites. These features are important to understand the mechanism of dynamics-driven allostery.

In the cases of GlmU and HCV NS5B.2, the importance of the ligand block size is suggested, as discussed in Results. In Table 5, we show overlap ratios of the predicted allosteric PLBs with the original allosteric ligands, which are defined as the ratio of the original ligand atoms covered by the PLBs. When the original allosteric ligands are employed, ΔMSF_{bs} value with the allosteric site model changed from 1 to 0.64 in GlmU and from 0.61 to 0.23 in HCV NS5B.2. PLB of GlmU overlaps with a part of the actual allosteric ligand (PDB ID: 2VD4 [29]) with the ratio of 0.24 (Fig. 3A). PLB of HCV NS5B.2 has more overlap with the actual allosteric ligand (0.42) (PDB ID: 2BRK [30]), but the key functional groups of the ligand are not generated (Fig. 3B). To overcome this problem, further refinement of the block generation method should be developed in the future. The overlap
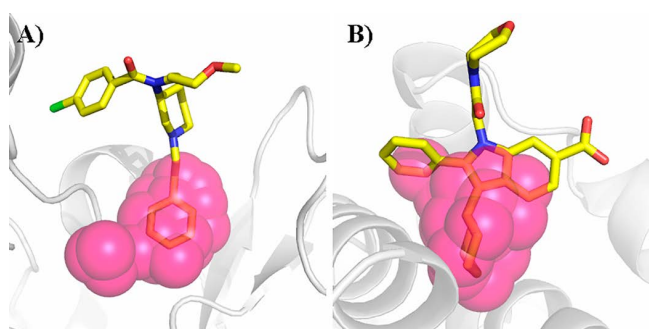
Figure 3 Size difference between PLBs and the allosteric ligands in the co-crystal. PLBs bound to the allosteric site (pink CPK) in A) GlmU (2VD4) and B) HCV NS5B.2 (2BRK), which are smaller than the allosteric ligands in the PDB (stick model).

ratio of CK2 and TEM1 is also low but the rank of ΔMSF_{bs} is relatively high. In these cases, the PLBs of the allosteric site models occupy the core parts of the pockets, which can significantly affect ΔMSF_{bs} values with relatively small blocks.

Among the 10 protein structures examined, large conformational change upon allosteric ligand binding was observed only in one case (glucokinase). However, the glucokinase complex with the active site ligand takes the structure very similar to the complex with the allosteric ligand as shown in Table 1. Small conformational changes at loop regions were observed in 3 cases (HCV NS5B.2, p38 MAP kinase.1, and TEM1). For more effective prediction, the effect of conformational change should be taken into account. To consider this, loop sampling methods should be combined with our method to conduct better generation of complex models. Another possibility is the use of the linear response theory [31] to predict large conformational changes such as hinge or domain motions. ΔMSF_{bs} did not work well in the case of the receptor protein; only one case in this category was found. In this case, the distance between the active and allosteric sites (r_{as-PLB}) is relatively short and MSF_{bs} is very small, which implies a different allosteric mechanism.

In addition to ΔMSF_{bs}, MSF_{bs}, and ζ are considered to be relatively good indicators although they are not always successful in distinguishing the allosteric sited from the other pockets. Nevertheless, ΔMSF_{bs}, MSF_{bs}, and ζ are suggested to be examined to find possible dynamic-driven allosteric sites. The results strongly suggest that dynamics is essential factor to characterize the allosteric effects of the proteins studied in this work.

Conclusion

The dynamic features of new types of allosteric sites were examined and quantities to characterize potential allosteric sites was developed based on the change of protein fluctuation upon ligand binding using ENM analysis. We focused on the non-trivial allosteric binding sites, assessed the static

and dynamic profiles, and found that fluctuation around the ligand binding site (ΔMSF_{bs}) was significantly suppressed upon binding to the allosteric site in 8 structures out of 9 enzymes examined in this work. One exception is the case in which the allosteric site occupies a part of the active site. These allosteric sites can be considered to be structurally sensitive to perturbation, inducing significant changes in the protein dynamics upon ligand binding. These cases should be regarded as examples of dynamics-driven allostery. Despite our initial speculation, there was generally no strong communication determined between the active and allosteric sites in the cases we examined.

ΔMSF_{bs} can be employed as a new type of the index to evaluate potential allosteric sites. Although ΔMSF_{bs} does not cover all types of allosteric sites, it can be used to search possible dynamics-driven allosteric sites. ΔMSF_{bs} is a relatively simple concept, where the potential allosteric sites of enzymes that are distant from the active site can be selected. This quantity may also be useful to find new types of drug target sites and lead to compound discovery. To improve the prediction power, it is essential to refine the block generation method and the prediction of possible conformational changes upon ligand binding.

The druggability is another important profile to be examined. The druggability of the dynamics-driven allosteric sites is significantly higher than that of the active sites and the others. The combination of ΔMSF_{bs} and the druggability can be considered as a good indicator to predict potential dynamics-driven allosteric sites.

Conflicts of Interest

J. T. and A. K. declare that they have no conflict of interest.

Author Contributions

J. T. and A. K. directed the entire project and co-wrote the manuscript. J. T. carried out calculations.

References

[1] Hardy, J. A. & Wells, J. A. Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* **14**, 706–715 (2004).

[2] Ming, D., Cohn, J. & Wall, M. Fast dynamics perturbation analysis for prediction of protein functional sites. *BMC Struct. Biol.* **8**, 5 (2008).

[3] Ming, D. & Wall, M. E. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* **95**, 198103 (2005).

[4] Ming, D. & Wall, M. E. Quantifying allosteric effects in proteins. *Proteins: Struct. Funct. Bioinform.* **59**, 697–707 (2005).

[5] Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).

[6] Ming, D. M. & Wall, M. E. Interactions in native binding sites cause a large change in protein dynamics. *J. Mol. Biol.* **358**, 213–223 (2006).

[7] Demerdash, O. N. A., Daily, M. D. & Mitchell, J. C. Structure-based predictive models for allosteric hot spots. *PLoS Comput. Biol.* **5**, e1000531 (2009).

[8] Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).

[9] Koshland, D. E., Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry (Mosc)* **5**, 365–385 (1966).

[10] Tsai, C.-J., del Sol, A. & Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **378**, 1–11 (2008).

[11] Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct. Funct. Bioinform.* **57**, 433–443 (2004).

[12] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., *et al.* The protein data bank. *Nucl. Acids Res.* **28**, 235–242 (2000).

[13] Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908 (1996).

[14] Scitegic, Pipeline Pilot [6.1.5] (2008).

[15] Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **46**, 3–26 (2001).

[16] Perry, J. J. P., Harris, R. M., Moiani, D., Olson, A. J. & Tainer, J. A. p38α MAP kinase C-terminal domain binding pocket characterized by crystallographic and computational analyses. *J. Mol. Biol.* **391**, 1–11 (2009).

[17] Pargellis, C., Tong, L., Churchill, L., Cirillo, P. F., Gilmore, T., Graham, A. G., *et al.* Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Mol. Biol.* **9**, 268–272 (2002).

[18] Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **80**, 505–515 (2001).

[19] Chennubhotla, C., Rader, A. J., Yang, L. W. & Bahar, I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* **2**, S173–180 (2005).

[20] Brady, G. P. & Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput.-Aided Mol. Des.* **14**, 383–401 (2000).

[21] Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **50**, 178–185 (1994).

[22] Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).

[23] Hayward, S., Kitao, A. & Go, N. Harmonicity and anharmonicity in protein dynamics—a normal-mode analysis and principal component analysis. *Proteins* **23**, 177–186 (1995).

[24] Kitao, A., Hayward, S. & Go, N. Energy landscape of a native protein: jumping-among-minima model. *Proteins* **33**, 496–517 (1998).

[25] Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L. & Camproux, A. C. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res.* **43**, W436–W442 (2015).

[26] MOE 2010.10, Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7 (2010).

[27] Yang, L. W. & Bahar, I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure* **13**, 893–904 (2005).

[28] Yang, L. W., Eyal, E., Bahar, I. & Kitao, A. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics* **25**, 606–614 (2009).

[29] Igor, M., Sandra, L., Lakshmi, N., Dirk, B., Michael, M., Steven, V., *et al.* Structure of a small-molecule inhibitor complexed with GlmU from *Haemophilus influenzae* reveals an allosteric binding site. *Protein Sci.* **17**, 577–582 (2008).

[30] Di Marco, S., Volpari, C., Tomei, L., Altamura, S., Harper, S., Narjes, F., *et al.* Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **280**, 29765–29770 (2005).

[31] Ikeguchi, M., Ueno, J., Sato, M. & Kidera, A. Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett.* **94**, 078102 (2005).

[32] Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J.-i. & Nagata, Y. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* **12**, 429 (2004).

[33] Yan, S., Appleby, T., Larson, G., Wu, J. Z., Hamatake, R., Hong, Z., *et al.* Structure-based design of a novel thiazolone scaffold as HCV NS5B polymerase allosteric inhibitors. *Bioorg. Med. Chem. Lett.* **16**, 5888 (2006).

[34] Raaf, J., Brunstein, E., Issinger, O.-G. & Niefind, K. The CK2α/CK2β interface of human protein kinase CK2 harbors a binding pocket for small molecules. *Chem. Biol.* **15**, 111 (2008).

[35] Horn, J. R. & Shoichet, B. K. Allosteric inhibition through core disruption. *J. Mol. Biol.* **336**, 1283–1291 (2004).

[36] Wiesmann, C., Barr, K. J., Kung, J., Zhu, J., Erlanson, D. A., Shen, W., *et al.* Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **11**, 730–737 (2004).

[37] Estébanez-Perpiñá, E., Arnold, L. A., Nguyen, P., Rodrigues, E. D., Mar, E., Bateman, R., *et al.* A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc. Natl. Acad. Sci. USA* **104**, 16074–16079 (2007).