Are the Apo Proteins Suitable for the Rational Discovery of Allosteric Drugs

An, X., Lu, S., Song, K., Shen, Q., Huang, M., Yao, X., Liu, H., & Zhang, J. (2018). Are the Apo Proteins Suitable for the Rational Discovery of Allosteric Drugs. *Journal of Chemical Information and Modeling*, 1-30. Advance online publication. https://doi.org/10.1021/acs.jcim.8b00735

Published in:
Journal of Chemical Information and Modeling

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Are the Apo Proteins Suitable for the Rational Discovery of Allosteric Drugs?

Journal:	*Journal of Chemical Information and Modeling*
Manuscript ID	ci-2018-007355.R1
Manuscript Type:	Article
Date Submitted by the Author:	27-Nov-2018

Complete List of Authors:	An, Xiaoli; Shanghai Jiao Tong University Lu, Shaoyong; Shanghai Jiao Tong University, Song, Kun; Shanghai Jiao Tong University School of Medicine, Department of Pathophysiology Shen, Qiancheng; Shanghai Jiao Tong University Huang, Meilan; Queen's University Belfast, School of Chemistry and Chemical Engineering Yao, Xiaojun; Lanzhou University, Chemistry Liu, Huanxiang; Lanzhou University, School of Pharmarcy Zhang, Jian; Shanghai Jiao Tong University,

Are the Apo Proteins Suitable for the Rational Discovery of Allosteric Drugs?

Xiaoli An[†,‡,#], Shaoyong Lu[†,#], Kun Song[†], Qiancheng Shen[†], Meilan Huang[¶], Xiaojun Yao[§], Huanxiang Liu[‡,*], Jian Zhang[†,∥,*]

[†]Key Laboratory of Cell Differentiation and Apoptosis of Ministry of Education, Clinical and Fundamental Research Center, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[‡]School of Pharmacy, Lanzhou University, Lanzhou 730000, China
[¶]School of Chemistry and Chemical Engineering, Queen's University Belfast, Northern Ireland BT9 5AG, United Kingdom
[§]State Key Laboratory of Quality Research in Chinese Medicine, Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Taipa, Macau 999078, China
[∥]Medicinal Bioinformatics Center, Shanghai Jiao Tong University, School of Medicine, Shanghai, 200025, China

[#] These authors contributed equally.

*Correspondence to: Jian Zhang (Phone: +86-21-63846590-776922; Fax: +86-21-64154900; E-mail: jian.zhang@sjtu.edu.cn)

*Correspondence to: Huanxiang Liu (Phone: +86-931-8915686; Fax: +86-931-8915685; E-mail: hxliu@lzu.edu.cn)

ABSTRACT

Allosteric modulators, by targeting the less-conserved allosteric sites, represent an innovative strategy in drug discovery. These modulators have a distinctive advantage over orthosteric ligands that attach to the conserved, functional orthosteric sites. However, in structure-based drug design, it remains unclear whether allosteric protein structures determined without orthosteric ligands binding are suitable for allosteric drug screening. In this study, we performed large-scale conformational samplings of six representative allosteric proteins uncomplexed (*apo*) and complexed (*holo*) with orthosteric ligands to explore the effect of orthosteric site binding on the conformational dynamics of allosteric sites. The results, coupled with the redocking evaluation of allosteric modulators to their *apo* and *holo* proteins using their MD trajectories, indicated that orthosteric site binding had an effect on the dynamics of the allosteric sites and allosteric modulators preferentially bound to their *holo* proteins. According to the analysis data, we constructed a new correlation model for quantifying the allosteric site change driven by substrate binding to the orthosteric site. These results highlight the strong demand to select *holo* allosteric proteins as initial inputs in structure-based allosteric drug screening when the distance between orthosteric and allosteric sites in the protein is below 5 Å, which is expected to contribute to allosteric drug discovery.

INTRODUCTION

Allostery, or allosteric regulation, regarded as ′the second secret of life′, fine-tunes most biological processes and controls physiological activities[1,2]. Dysregulation of protein allostery can lead to human diseases; therefore, recovery of malfunctional proteins to their normal functions by allosteric modulators provides a rich landscape for new therapeutics[3,4].

Allosteric modulators, by targeting the less conserved allosteric sites that are topologically and spatially distinct from the conserved, functional, orthosteric sites, bear distinctive advantages compared to orthosteric ligands, including higher specificity and lower toxicity[5-7]. Despite a number of potential advantages endowed by allosteric therapeutics, in recent years, the discovery allosteric drugs has been challenging. Indeed, a vast majority of allosteric modulators have been identified serendipitously by means of high-throughput screening[8]. This difficulty has severely hindered progress in the development of allosteric drugs over the past several decades, as demonstrated by the limited number of allosteric modulators reported to date[9,10].

Due to recent advances in X-ray and NMR spectroscopies, a considerable number of allosteric proteins complexed or uncomplexed with their allosteric modulators have been discovered[11-13]. With the characterization of allosteric proteins and allosteric protein-modulator complexes, the unique hallmarks of allosteric proteins, allosteric modulators, and allosteric sites have been increasingly characterized[8,14-16]. The increased availability of structural data, coupled with the advances in computer power, facilitates the development of structure-based computational methods for detection of

allosteric sites in proteins because identification of allosteric sites is the first step for *in silico* screening of potential allosteric modulators. To date, a series of structure-based predictive models have been developed to predict allosteric sites and deployed to web servers, such as Allosite[17], PARS[18], and SPACER[19].

A prerequisite for the success of *in silico* screening is to provide an input structure that can represent a major physiological conformation in the cellular environment. As in the case of protein allostery, binding of an allosteric modulator to its allosteric site alters the shape and dynamic of the orthosteric site, suggesting that there is signal propagation from allosteric to orthosteric sites[20-22]. Consistently, biophysical experiments indicated that orthosteric site occupancy by its endogenous substrate or exogenous ligands can also alter the conformational dynamics of allosteric site, reflecting the bidirectional regulation between allosteric and orthosteric sites[23,24]. When the known allosteric proteins deposited into the Allosteric Database (ASD)[9] are under intense scrutiny, a number of allosteric proteins are crystallized without the endogenous substrates or exogenous ligands bound to their orthosteric sites (hereafter called *apo* proteins). This observation raises a significant question: are the *apo* (orthosteric site-unbound) protein structures suitable as initial conformations for *in silico* allosteric screening?

To address this question, we sampled large-scale conformations of the *apo* and *holo* (orthosteric site-bound) proteins to quantify the effect of orthosteric site binding on the conformational dynamics of the allosteric site. Six representative allosteric proteins were selected, including tryptophanyl-tRNA synthetase (TrpRS), human monoamine

oxidase B (MAO B), N-acetylglucosamine-1-phosphate uridyltransferase (GlmU), macrophage migration inhibitory factor (MIF), ERK/MAP kinase (MAPK), and GTPase K-Ras4B oncoprotein. The analysis results revealed that the conformational dynamics of allosteric sites changed significantly in response to orthosteric site binding. Furthermore, redocking of allosteric modulators to the *apo* and *holo* proteins extracted from the MD trajectories indicated that allosteric modulators preferentially bound to their holo proteins. Finally, we constructed a new correlation model for evaluating the allosteric site change driven by substrate binding into the orthosteric site. Taken together, these results support the use of the *holo* allosteric proteins as initial inputs in structure-based drug screening, which is expected to contribute to allosteric drug discovery.

METHODS

Dataset

First, we manually picked allosteric proteins from our constructed ASD[9] in which the orthosteric site is adjacent to the allosteric site. Next, we examined the residues located in orthosteric and allosteric sites and ruled out the proteins without shared residues in the orthosteric and allosteric sites. The final selected proteins consisted of six enzymes: TrpRS (PDB ID 1MAU)[25], ERK/MAP kinase (MAPK) (PDB ID 4ANB)[26], K-Ras4B (PDB ID 4LUC)[27], GlmU (PDB ID 2V0I)[28], MAO B (PDB ID 2XCG)[29], and MIF (PDB ID 3IJJ)[30] (Figure 1). Moreover, the phylogenetic tree constructed for the six allosteric proteins displayed a wide distribution of these allosteric proteins, indicating that the allosteric proteins belong to different protein families.

Preparation of simulated systems

In all six allosteric proteins, the *holo* forms were obtained by removing the allosteric modulators from the allosteric sites. The *apo* forms were obtained by removing both allosteric modulators and orthosteric ligands from the allosteric and orthosteric sites. The chain A was selected for the dimer of MAO B and K-Ras4B. The solvent molecules except crystal water were deleted and the metal ions in pockets were kept. In the MAO B, there has a flavin cofactor that was reserved in both holo and apo forms during simulations. In each protein, a pair of face-to-face residues located in the middle of the tunnel between the allosteric and orthosteric sites were defined as bayonet residues. In addition, the orthosteric and allosteric site residues were annotated. Special case is the crystal structure of GlmU (PDB ID 2V0I)[28], because there is a lack of an allosteric ligand. However, the crystal structure of GlmU in complex with an allosteric inhibitor (PDB ID 2VD4)[31] reveals that allosteric site is adjacent to orthosteric site. We thus aligned the two crystal structures to monitor the minimum distance between the two ligands and define the bayonet residues.

The proteins were modeled using the AMBER FF99SB force field[32], while the orthosteric ligands were modeled using the generalized AMBER force field (GAFF)[33]. Geometry optimization and the electrostatic potential calculations on the allosteric modulators or orthosteric ligands were performed at the HF/6-31G* level of Gaussian09[34], and the partial charges were calculated with the restrained electrostatic potential (RESP) fitting method[35]. The force field parameters for the allosteric modulators or orthosteric ligands were created by the Antechamber package. For the metal atoms, GTP and ATP were obtained from the AMBER parameter database[36]. Each system was dissolved in a truncated octahedral box of TIP3P water[37] and was neutralized with counter ions.

MD simulations

MD simulations for each system were performed using the AMBER11 package[38-41]. First, to remove bad contacts in the initial structures, steepest descent and conjugate gradient algorithms were performed for energy minimization. After energy minimization, each system was gradually heated from 0 to 300 K in 100 ps. Subsequently, constant temperature equilibration at 300 K for 300 ps was performed to adjusting the solvent density. Finally, 200-ns MD simulations were carried out for each system in the NPT ensemble with periodic boundary conditions. An integration step of 2 fs was used, and the long-range electrostatic interactions were treated by the particle mesh Ewald (PME) method[42]. A cut-off of 10 Å was used for limiting the direct space interactions. The SHAKE method[43] was applied to constrain all covalent bonds involving hydrogen atoms. Each simulation was coupled to at a temperature of 300 K and a pressure of 1.0 atm by applying the Langevin algorithm[44].

Dynamical cross-correlation matrices

Dynamic cross-correlation matrices (DCCM) were used to detect time-correlation motion in protein, which is composed of the fluctuation cross-correlations coefficient in the positions of Cα atoms during the MD simulation.[45] The normalized cross-correlation function C_{ij} was calculated according to Eq. (1):

$$C_{ij} = \frac{\langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \rangle}{\left(\langle \Delta\vec{r}_i(t) \rangle^2 \langle \Delta\vec{r}_j(t) \rangle^2 \right)^{1/2}} \quad (1)$$

$$\Delta\vec{r}_i(t) = \vec{r}_i(t) - \langle \vec{r}_i(t) \rangle \quad (2)$$

where $\vec{r}_i(t)$ and $\vec{r}_j(t)$ are the spatial positions of Cα atoms corresponding residues *i* and *j* at time *t*. The value of C_{ij} is between -1~1. $C_{ij} > 0$ indicated the Cα atoms between residues *i* and *j* exhibited motions in same direction along a given spatial coordinate, whereas $C_{ij} < 0$ indicated the Cα atoms between residues *i* and *j* exhibited

motions in opposite directions along a given spatial coordinate. $C_{ij} = 0$ indicated the Cα atoms exhibited motions between residues *i* and *j* independently. The correlations were calculated by the *ptraj* module in AMBER.

RESULTS AND DISCUSSION

Correlation of allosteric site volume and bayonet residues

To test our assumption that binding to the orthosteric site has an effect on the dynamics of allosteric site, we first monitored potential dynamic changes between two bayonet residues using 5000 frames from the last 100 ns trajectories. Next, the allosteric pocket volumes of the *apo* and *holo* forms, which have a cavity shape, were calculated using MDpocket[46]. The results for all systems are displayed in the scatter diagram to intuitively show distance and volume (lnV) variations (Figure 2). In TrpRS (Figure 2A), ERK/MAP kinase (MAPK) (Figure 2B), K-Ras4B (Figure 2C), and GlmU (Figure 2D), both the allosteric site volume and bayonet residue distance increased significantly in the *holo* forms compared to the *apo* forms. However, in MAO B (Figure 2E), no appreciable differences in the allosteric site volume or bayonet residue distance between the *holo* and *apo* forms were observed. This exception is partly due to the evidence that the allosteric site is relatively far from the active site (Figure 1), thereby leading to the subtle effect of orthosteric site binding on the dynamics of allosteric site for our timescale. Unexpectedly, in MIF (Figure 2F), both the allosteric site volume and bayonet residue distance decreased significantly in response to orthosteric site binding. Overall, regardless of the enlargement or contraction of the allosteric sites, the

topological structures of the allosteric sites changed remarkably in reaction to orthosteric site binding.

Concerted atomic motions of orthosteric and allosteric sites

Given the influence of orthosteric site binding on the conformational dynamics of the allosteric site, we next analyzed DCCM using 5000 frames from the last 100 ns trajectories to reveal concerted atomic motions of orthosteric and allosteric sites in the *apo* and *holo* forms, respectively. In this analysis, we primarily focused on the regions formed by orthosteric and allosteric sites (the residues within 5 Å of ligands), as well as those between them, which are named a shared community. Furthermore, the differences in DCCM between the *apo* and *holo* forms were calculated to understand the altered motions of the allosteric site upon binding of the corresponding orthosteric ligand.

In TrpRS (Figure 3A), ERK/MAP kinase (MAPK) (Figure 3B), K-Ras4B (Figure 3C), and GlmU (Figure 4A), the correlated motions of allosteric sites, orthosetric sites, and shared community remarkably increased in the *holo* forms compared to the *apo* forms, strongly supporting the notion that orthosteric ligand binding to the orthosteric site affects the conformational dynamics of the corresponding allosteric site. In the four allosteric proteins, the allosteric sites are proximal to their orthosteric sites. As a result, the fluctuations in the orthosteric sites induced by orthosteric ligand binding can readily propagate to their nearby allosteric sites through the shared community. In MAO B (Figure 4B), the correlated motions of the allosteric site were unaffected in the presence of an allosteric ligand at the allosteric site because of the distant location of the allosteric

and orthosteric sites. This result is consistent with the analysis of allosteric site volume and bayonet residue distance in MAO B. In MIF (Figure 4C), the correlated motions of the allosteric site in the *holo* form slightly increased in relation to the *apo* form. Taken together, the results of the DCCM analysis demonstrate that binding to the orthosteric site has the potential to alter the dynamics of the allosteric site.

Evaluation of allosteric molecular binding mode

The assessment of the allosteric site volume and the coupled motions of the allosteric site suggest the variation of the allosteric site's topological structure as a consequence of orthosteric site binding. To further verify the significance of conformational variation in the allosteric sites of the *holo* forms, redocking of allosteric modulators to their corresponding allosteric sites in both the *apo* and *holo* forms was performed using Glide[47]. First, the hierarchical clustering analysis was performed with the last 100 ns MD trajectories (10000 frames) of the apo and holo forms, and each system was grouped into five clusters. Representative structure was extracted from each cluster of the *apo* and *holo* forms of each system, respectively. The five representative structures obtained after cluster analysis can represent a wide range of different conformational states. Next, redocking of allosteric modulators to the allosteric sites of the five respective structures in both the *apo* and *holo* forms was performed, and 32 docked poses were obtained for each allosteric modulator to the allosteric sites of *apo* and *holo* forms, respectively. Herein, we only considered the re-docked binding poses of allosteric ligands. Therefore, the root-mean-squared-deviation (RMSD) of each allosteric modulator between the native binding mode and the re-docked binding mode

was calculated, and the best result of each re-docked complex was used as an evaluation criteria of the redocking binding mode.

In TrpRS (Figure 5A), ERK/MAP kinase (MAPK) (Figure 5B), K-Ras4B (Figure 5C), and GlmU (Figure 5D), the RMSD of the allosteric modulators docked into the *holo* forms was significantly smaller than those docked into the *apo* forms. In MAO B (Figure 5E), docked conformations of the allosteric modulator to the allosteric sites of the *apo* and *holo* forms had no remarkable differences, reflecting no significant effect of orthosteric site binding on the topology of the allosteric site for our timescale. In MIF (Figure 5F), the RMSD of the allosteric modulator to the *holo* form was slightly smaller than that to the *apo* form.

The redocking results suggest a difference in the geometrical conformation of allosteric sites in the presence and absence of orthosteric site binding. Moreover, the allosteric modulators can fit snugly into their allosteric sites in the presence of orthosteric ligands, highlighting the significance of orthosteric site binding in re-shaping the allosteric site. This principle can be applied for structure-based allosteric molecule screening against the crystal structures of orthosteric ligand-bound proteins.

Correlation of minimum distance between allosteric and orthosteric sites with allosteric site volume

The analyses of allosteric site volume, DCCM, and redocking experiments revealed that when allosteric sites are in close proximity to orthosetric sites, the conformational variations in allosteric sites are apparent upon the binding of the orthosteric site. Thus, we hypothesize that the extent of allosteric site conformational

variations induced by the binding of the orthosteric site is dependent on the distance between the allosteric and orthosteric sites. To validate this hypothesis, we correlated the distance between allosteric and orthosteric sites with the perturbation of allosteric site upon orthosteric site binding. The centroid distance between allosteric and orthosteric sites was used as the location of the two sites, and the allosteric site volume was used to evaluate the allosteric site conformational variation between the *apo* and *holo* forms. The fitted curve of distances (d) and volume variations ($\Delta\overline{V}$) are shown in Figure 6. The fitted curve can be formulated as

$$y = \exp(a + bx) \quad (3)$$

where y is the $\Delta\overline{V}$ of the allosteric site between the *holo* and *apo* forms, and x is the d between the allosteric and orthosteric sites. The fitted constants *a* and *b* are 9.93 and -1.17, respectively. Goodness of fit was evaluated with the R-square (R^2), where $R^2 = 0.924$.

Based on the curve, it is noticeable that if the distance between allosteric and orthosteric site is less than 5 Å, the $\Delta\overline{V}$ of the allosteric site alters significantly with the variation of distance. This finding indicates that when the allosteric site is close to its orthosteric site, the conformational dynamics of the allosteric site will be markedly changed after orthosteric site binding. However, if the distance between allosteric and orthosteric sites is larger than 5 Å, the variation of distance has a minor effect on the $\Delta\overline{V}$ of the allosteric site.

CONCLUSIONS

Due to the prominent functional importance of allostery in controlling physiological activities and causing human diseases, allosteric drug development is an innovative research strategy. Allosteric drugs do not compete with endogenous substrates that occupy the conserved orthosteric sites. In contrast, they attach to the structurally diverse allosteric sites, thereby possessing higher selectivity compared to orthosteric drugs. However, the design of allosteric drugs is currently challenging because unearthing the allosteric sites in proteins is a challenge[4,48]. Indeed, the majority of currently identified and characterized allosteric sites were adventitiously discovered by biochemical experiments.[49-50] In addition, in structure-based allosteric drug design, the hit rate for allosteric screening is extremely low with notably few instances of allosteric modulator identification via computational screening[10]. There is a lack of knowledge regarding the interplay between allosteric and orthosteric sites.[51]

The recent unified view of allostery introduced by Nussinov and Tsai emphasizes the existence of structural coupling between the allosteric and orthosetric sites[3,52]. The allosterically perturbed signals transmit from allosteric to orthosetric sites, thereby affecting the functional activities of orthosteric sites. In a similar vein, using a Gaussian network model to analyze motion correlation of allosteric proteins, Ma *et al*.[53] revealed pronounced correlations between allosteric and orthosteric sites in both monomeric and oligomeric allosteric proteins. These studies are indicative of the bidirectional modulation between allosteric and orthosteric sites. In other words, allosteric site binding enables reshaping of the corresponding orthosteric site and vice versa.

In the allosteric drug design, the appropriate selection of the allosteric site in the

three-dimensional structure becomes a high-priority goal, which determines the success of allosteric modulator identification. In effect, a considerable proportion of allosteric proteins are discovered without the presence of endogenous substrates or orthosteric ligands bound to their orthosteric sites. To definitively elucidate the effect of orthosteric site binding on the conformational dynamics of allosteric sites, we performed MD simulations of six representative allosteric proteins in both *apo* and *holo* forms. Through the analyses of allosteric site volume and correlated motions of allosteric and orthosteric sites, we observed that when the allosteric and orthosteric sites are closely positioned on the protein, the conformational plasticity of the allosteric site is markedly affected by binding of the orthosteric ligand to the orthosteric site. As a result, redocking experiments corroborated that the allosteric modulators prefer binding to their corresponding allosteric proteins in the presence of a bound orthosteric site. However, the correlation model indicates that the reverse regulation between allosteric site and orthosteric site may depend on the distance between the two sites. Most likely, when the distance between the orthosteric and allosteric sites in the protein is below 5 Å, it is advisable to select the crystal structures of allosteric proteins with their orthosteric ligands bound as the initial inputs to structure-based allosteric drugs screening.

ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China (Grant No. 21675070 and No. 21778037) and the Fundamental Research Funds for the Central

Universities (Grant No. lzujbky-2017-k24).

References

(1) Fenton, A. W. Allostery: an Illustrated Definition for the Second Secret of Life. *Trends Biochem. Sci.* **2008**, *33*, 420-425.

(2) Lindsley, J. E.; Rutter J. Whence Cometh the Allosterome? *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10533-10535.

(3) Nussinov, R.; Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153*, 293-305.

(4) Lu, S.; Li, S.; Zhang, J. Harnessing Allostery: a Novel Approach to Drug Discovery. *Med. Mes. Res.* **2014**, *6*, 1242-1285.

(5) Goodey, N. M.; Benkovic, S. J. Allosteric Regulation and Catalysis Emerge via a Common Route. *Nat. Chem. Biol.* **2008**, *4*, 474-482.

(6) Wootten, D.; Christopoulos, A.; Sexton, P. M. Emerging Paradigms in GPCR Allostery: Implications for Drug Discovery. *Nat. Rev. Drug Discov.* **2013**, *12*, 630-644.

(7) Lu, S.; Zhang, J. Designed Covalent Allosteric Modulators: an Emerging Paradigm in Drug Discovery. *Drug Discov. Toady* **2017**, *22*, 447-453.

(8) Lu, S.; Huang, W.; Zhang, J. Recent Computational Advances in the Identification of Allosteric Sites in Proteins. *Drug Discov. Today* **2014**, *19*, 1595-1600.

(9) Shen, Q.; Wang, G.; Li, S.; Liu, X.; Lu, S.; Chen, Z.; Song, K.; Yan, J.; Geng, L.; Huang, Z.; Huang, W.; Chen, G.; Zhang, J. ASD v3.0: Unraveling Allosteric Regulation with Structural Mechanisms and Biological Networks. *Nucleic Acids Res.* **2016**, *44*, D527-D535.

(10) Lu, S.; Zhang, J. Allosteric Modulators. *Comprehen. Med. Chem. III* **2017**, *2*, 276-296.

(11) Huang, W.; Wang, G.; Shen, Q.; Liu, X.; Lu, S.; Geng, L.; Huang, Z.; Zhang, J. ASBench: Benchmarking Sets for Allosteric Discovery. *Bioinformatics* **2015**, *31*, 2598-2600.

(12) Boulton, S.; Melacini, G. Advances in NMR Methods to Map Allosteric Sites: From Models to Translation. *Chem. Rev.* **2016**, *116*, 6267-6304.

(13) Lu, S.; Zhang, J. Small Molecule Allosteric Modulators of G‐protein-Coupled Receptors: Drug−Target Interactions. *J. Med. Chem.* 2018, doi: *10.1021/acs.jmedchem.7b01844*.

(14) Li, X.; Chen, Y.; Lu, S.; Huang, Z.; Liu, X.; Wang, Q.; Shi, T.; Zhang, J. Toward an Understanding of the Sequence and Structural Basis of Allosteric Proteins. *J. Mol. Graph Modell* **2013**, *40*, 30-39.

(15) van Westen, G. J. P.; Gaulton, A.; Overington, J. P. Chemical, Target, and Bioactive Properties of Allosteric Modulation. *PLoS Comput. Biol.* **2014**, *10*, e1003559.

(16) Yang, J. S.; Seo S. W.; Jang, S.; Jung, G. Y.; Kim, S. Rational Engineering of Enzyme Allosteric Regulation through Evolution Analysis. *PLoS Comput. Biol.* **2012**, *8*, e1002612.

(17) Huang, W.; Lu, S.; Huang, Z.; Liu, X.; Mou, L.; Luo, Y.; Zhao, Y.; Liu, Y.; Chen, Z.; Hou, T.; Zhang, J. Allosite: a Method for Predicting Allosteric Sites. *Bioinformatics* **2013**, *29*, 2357-2359.

(18) Panjkovich, A.; Daura, X. PARS: a Web Server for the Prediction of Protein Allosteric and Regulatory Sites. *Bioinformatics* **2014**, *30*, 1314-1315.

(19) Goncearenco, A.; Mitternacht S.; Yong T.; Eisenhaber B.; Eisenhaber F.; Berezovsky I. N. SPACER: Server for Predicting Allosteric Communication and Effects of Regulation. *Nucleic Acids Res.* **2013**, *41*, W266-W272.

(20) Guarnera, E.; Berezovsky, I. N. Allosteric Sites: Remote Control in Regulation of Protein Activity. *Curr. Opin. Struct. Biol.* **2016**, *37*, 1-8.

(21) Pei, J.; Yin, N.; Ma, X.; Lai, L. Systems Biology Brings New Dimensions for Structure-Based Drug Design. *J. Am. Chem. Soc.* **2014**, *136*, 11556-11565.

(22) del Sol, A.; Tsai, C. J.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure* **2009**, *17*, 1042-1050.

(23) Guo, J.; Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503-6515.

(24) Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J. Ras Conformational Ensembles, Allostery, and Signaling. *Chem. Rev.* **2016**, *116*, 6607-6665.

(25) Retailleau, P.; Huang, X.; Yin, Y.; Hu, M.; Weinreb, V.; Vachette, P.; Vonrhein, C.; Bricogne, G.; Roversi, P.; Ilyin, V.; Carter, C. W. Jr. Interconversion of ATP Binding and Conformational Free Energies by Tryptophanyl-tRNA Synthetase: Structures of ATP Bound to Open and Closed, Pre-transition-state Conformations. *J. Mol. Biol.* **2003**, *325*, 39-63.

(26) Rice, K. D.; Aay, N.; Anand, N. K.; Blazey C. M.; Bowles, O. J.; Bussenius, J.; Costanzo, S.; Curtis, J. K.; Defina, S. C.; Dubenko, L.; Engst, S.; Joshi, A. A.; Kennedy, A. R.; Kim, A.I.; Koltun, E. S.; Lougheed, J. C.; Manalo, J. C.; Martini, J. F.; Nuss, J. M.; Peto, C. J.; Tsang, T. H.; Yu, P.; Johnston, S. Novel Carboxamide-based Allosteric MEK Inhibitors: Discovery and Optimization Efforts toward XL518 (GDC-0973). *ACS Med. Chem. Lett.* **2012**, *3*, 416-421.

(27) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells J. A.; Shokat K. M. K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **2013**, *503*, 548-551.

(28) Mochalkin, I.; Lightle, S.; Narasimhan, L.; Bornemeier, D.; Melnick, M.; Vanderroest, S.; McDowell, L. Structure of a Small-molecule Inhibitor Complexed with GlmU from *haemophilus influenzae* Reveals an Allosteric Binding Site. *Protein Sci.* **2008**, *17*, 577-582.

(29) Bonivento, D.; Milczek, E. M.; McDonald, G. R.; Binda, C.; Holt, A.; Edmondson, D .E.; Mattevi, A. Potentiation of Ligand Binding through Cooperative Effects in Monoamine Oxidase B. *J. Biol. Chem.* **2010**, *285*, 36849-36856.

(30) Cho, Y.; Crichlow, G. V.; Vermeire, J. J.; Leng L.; Du X.; Hodsdon M. E.; Bucala R.; Cappello M.; Gross M.; Gaeta F.; Johnson K.; Lolis E. J. Allosteric Inhibition of Macrophage Migration Inhibitory Factor Revealed by Ibudilast. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11313-11318.

(31) Mochalkin, I.; Lightle, S.; Narasimhan, L.; Bornemeier, D.; Melnick, M.; Vanderroest, S.; McDowell, L. Structure of a Small-molecule Inhibitor Complexed

with GlmU from Haemophilus Influenzae Reveals an Allosteric Binding Site. *Protein Sci.* 2016, *17*, 577-582.

(32) Hornak, V.; Abel, R.; Okur, A.; Strockbine B.; Roitberg A. Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712-725.

(33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman P. A.; Case D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem*. **2004**, *25*, 1157-1174.

(34) Gaussian, version 09; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.: Inc., Wallingford CT. 2009.

(35) Fox, T.; Kollman, P. A. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem.* **1988**, *102*, 8070-8079.

(36) Homeyer, N.; Horn, A. H.; Lanig, H.; Sticht, H. AMBER Force-field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J. Mol. Model.* **2006**, *12*, 281-289.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(38) AMBER, version 11; Case, A.D.: San Francisco, 2010.

(39) Jang, H.; Banerjee, A.; Chavan, T.S.; Lu, S.; Zhang, J.; Gaponenko, V.; Nussinov, R. The Higher Level of Complexity of K-Ras4B Activation at the Membrane. *FASEB J*. **2016**, *30*, 1643-1655.

(40) Li, J.; Sun, L.; Xu, C.; Yu, F.; Zhou, H.; Zhao, Y.; Zhang, J.; Cai, J.; Mao, C.; Tang, L.; Xu, Y.; He, J. Structure Insights into Mechanisms of ATP Hydrolysis and the Activation of Human Heat-shock Protein 90. *Acta Biochim Biophys Sin (Shanghai)*. **2012**, *44*, 300-306.

(41) Han, C.; Zhang, J.; Chen, L.; Chen, K.; Shen, X.; Jiang, H. Discovery of Helicobacter Pylori Shikimate Kinase Inhibitors: Bioassay and Molecular Modeling. *Bioorg Med Chem*. **2007**, *15*, 656-662.

(42) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(43) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1997**, *23*, 321-341.

(44) Wu, X.; Brooks, B. R. Self-guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381*, 512-518.

(45) Chen, S.; Hu, T.; Zhang, J.; Chen, J.; Chen, K.; Ding, J.; Jiang, H.; Shen, X. Mutation of Gly-11 on the Dimer Interface Results in the Complete Crystallographic Dimer Dissociation of Severe Acute Respiratory Syndrome Coronavirus 3C-like Protease: Crystal Structure with Molecular Dynamics Simulations. *J Biol Chem*. **2008**, *283*, 554-564.

(46) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: Open-source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27*, 3276-3285.

(47) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750-1759.

(48) Lu, S.; Ji, M.; Ni, D.; Zhang, J. Discovery of Hidden Allosteric Sites as Novel Targets for Allosteric Drug Design. *Drug Discov. Today.* **2018**, *23*, 359-365.

(49) Huang, Z.; Zhao, J.; Deng, W.; Chen, Y.; Shang, J.; Song, K.; Zhang, L.; Wang, C.; Lu, S.; Yang, X.; He, B.; Min, J.; Hu, H.; Tan, M.; Xu, J.; Zhang, Q.; Zhong, J.; Sun, X.; Mao, Z.; Lin, H.; Xiao, M.; Chin, YE.; Jiang, H.; Xu, Y.; Chen, G.; Zhang, J. Identification of a Cellularly Active SIRT6 Allosteric Activator. *Nat Chem Biol*. **2018**, *14*, 1118-1126.

(50) Shen, Q.; Cheng, F.; Song, H.; Lu, W.; Zhao, J.; An, X.; Liu, M.; Chen, G.; Zhao, Z.; Zhang, J. Proteome-Scale Investigation of Protein Allosteric Regulation Perturbed by Somatic Mutations in 7,000 Cancer Genomes. *Am J Hum Genet.* **2017**, *100*, 5-20.

(51) Jiang, H.; Deng, R.; Yang, X.; Shang, J.; Lu, S.; Zhao, Y.; Song, K.; Liu, X.; Zhang, Q.; Chen, Y.; Chinn, Y.E.; Wu, G.; Li, J.; Chen, G.; Yu, J.; Zhang, J. Peptidomimetic Inhibitors of APC-Asef Interaction Block Colorectal Cancer Migration. *Nat. Chem. Biol*. **2017**, *13*, 994-1001.

(52) Tsai, C. J.; Nussinov, R. A Unified View of ''How Allostery Works''. *PLoS Comput. Biol.* **2014**, *10*, e1003394.

(53) Ma, X.; Meng, H.; Lai, L. Motions of Allosteric and Orthosteric Ligand-binding Sites in Proteins Are Highly Correlated. *J. Chem. Inf. Model.* **2016**, *56*, 1725-1733.

FIGURES

Figure 1. Phylogenetic tree analyses of six representative allosteric proteins, including tryptophanyl-tRNA synthetase (TrpRS) (PDB ID 1MAU), human monoamine oxidase B (MAO B) (PDB ID 2XCG), N-acetylglucosamine-1-phosphate uridyltransferase (GlmU) (PDB ID 2V0I), macrophage migration inhibitory factor (MIF) (PDB ID 3IJJ), ERK/MAP kinase (MAPK) (PDB ID 4ANB), and GTPase K-Ras4B (PDB ID 4LUC). Phylogenetic distance is proportional to branch length (the connecting lines). The allosteric and orthosteric sites in the proteins are extracted and depicted by splitpea and cyan, respectively. Allosteric modulators and orthosteric ligands in their corresponding allosteric and orthosteric sites are depicted by stick models.

Figure 2. Variations of the allosteric site volumes and bayonet residue distances in TrpRS (A), ERK/MAP kinase (MAPK) (B), K-Ras4B (C), and GlmU (D), MAO B (E), and MIF (F).

Figure 3. Dynamic cross-correlation maps (DCCM) reveal the extent of correlation for the orthosteric and allosteric sites as well as the shared community between them in the *apo* and *holo* forms of TrpRS (A), ERK/MAP kinase (MAPK) (B), and K-Ras4B (C). The absolute values for the differences in DCCM between the *apo* and *holo* forms are shown in the right panel. Motion occurring along the same direction is represented by positive correlation (orange), while anti-correlated motion occurring along the opposite direction is represented by negative correlation (green). ′S′ and ′A′ represent orthosteric and allosteric sites, respectively.

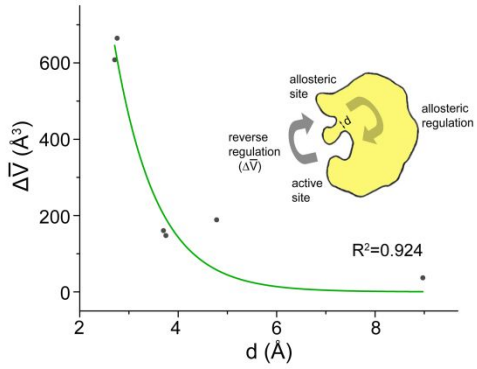
Figure 4. Dynamic cross-correlation maps (DCCM) reveal the extent of correlation for the orthosteric and allosteric sites as well as the shared community between them in the *apo* and *holo* forms of GlmU (A), MAO B (B), and MIF (C). The absolute values for the differences in DCCM between the *apo* and *holo* forms are shown in the right panel. Motion occurring along the same direction is represented by positive correlation

(orange), while anti-correlated motion occurring along the opposite direction is represented by negative correlation (green). ′S′ and ′A′ represent orthosteric and allosteric sites, respectively.

Figure 5. Root-mean-squared-deviation (RMSD) of each allosteric modulator between the native-binding mode and the re-docked binding mode for the five representative structures extracted from the MD trajectory. The blue and orange bars represent the RMSD for the *apo* and *holo* forms, respectively. TrpRS (A), ERK/MAP kinase (MAPK) (B), K-Ras4B (C), GlmU (D), MAO B (E), and MIF (F).

Figure 6. Relationship between the distance of the centroids between allosteric and orthosteric sites and the allosteric site volume.

Table of Contents Graphic



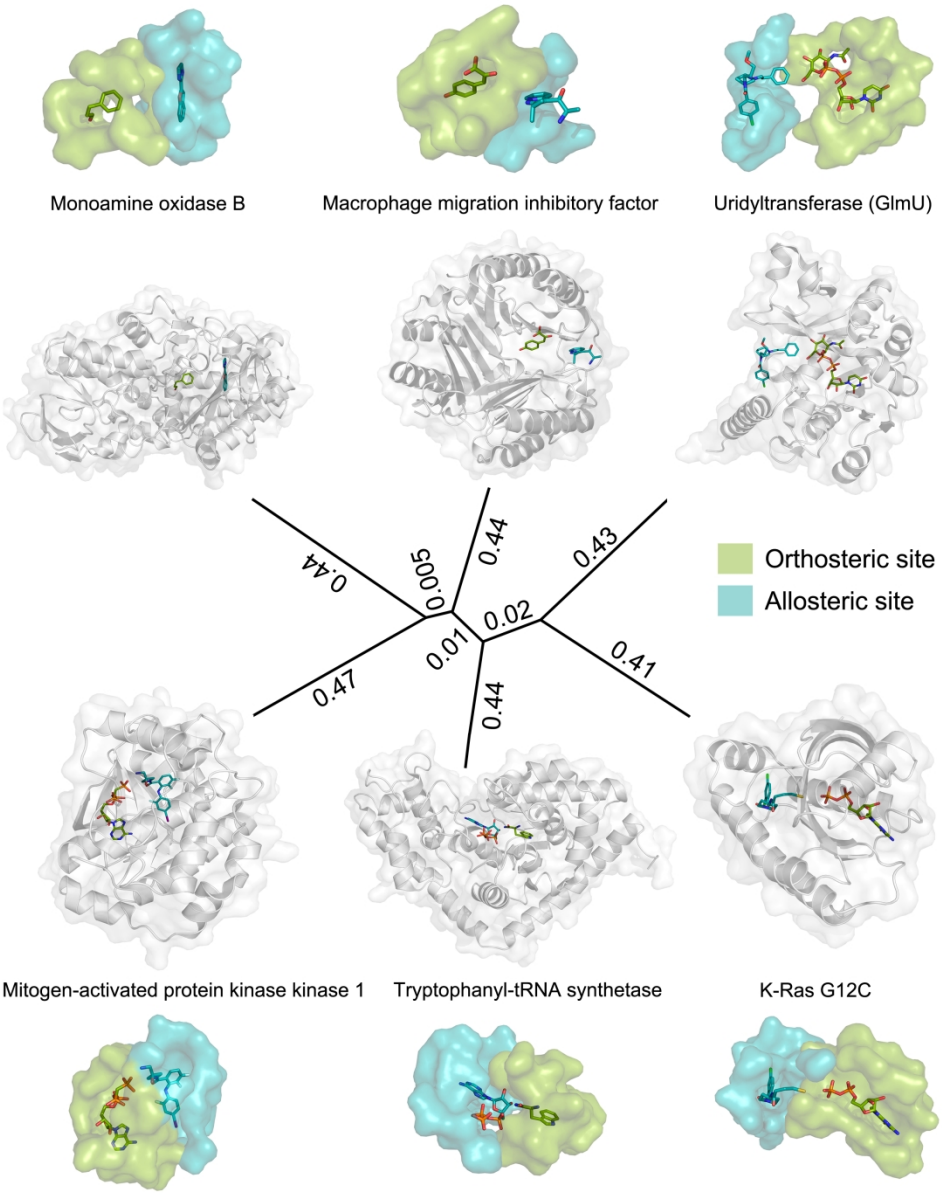


Figure 1. Phylogenetic tree analyses of six representative allosteric proteins, including tryptophanyl-tRNA synthetase (TrpRS) (PDB ID 1MAU), human monoamine oxidase B (MAO B) (PDB ID 2XCG), N-acetylglucosamine-1-phosphate uridyltransferase (GlmU) (PDB ID 2V0I), macrophage migration inhibitory factor (MIF) (PDB ID 3IJJ), ERK/MAP kinase (MAPK) (PDB ID 4ANB), and GTPase K-Ras4B (PDB ID 4LUC). Phylogenetic distance is proportional to branch length (the connecting lines). The allosteric and orthosteric sites in the proteins are extracted and depicted by splitpea and cyan, respectively. Allosteric modulators and orthosteric ligands in their corresponding allosteric and orthosteric sites are depicted by stick models.

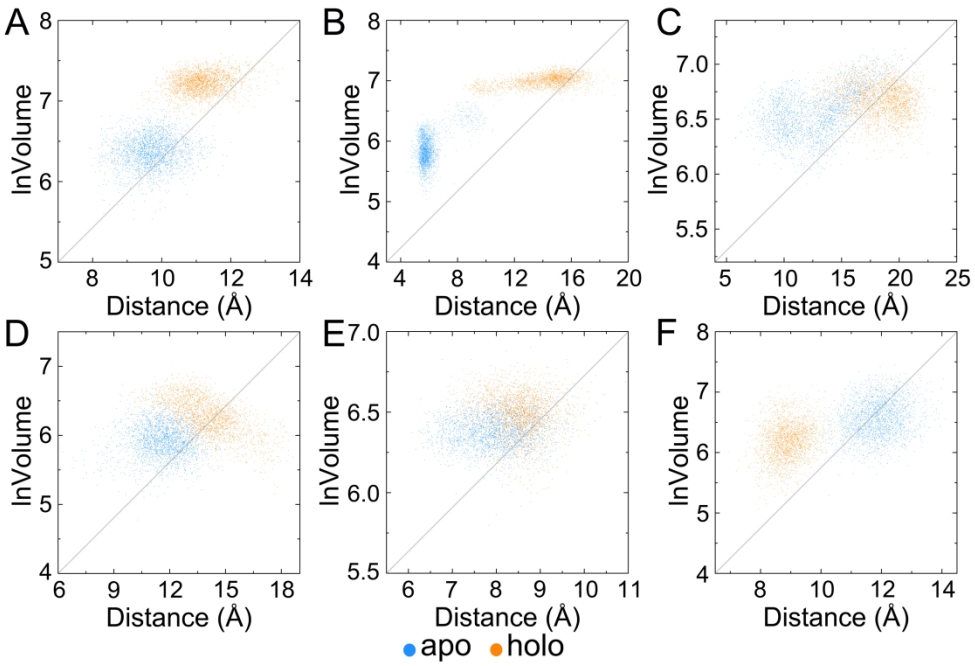


Figure 2. Variations of the allosteric site volumes and bayonet residue distances in TrpRS (A), ERK/MAP kinase (MAPK) (B), K-Ras4B (C), and GlmU (D), MAO B (E), and MIF (F).

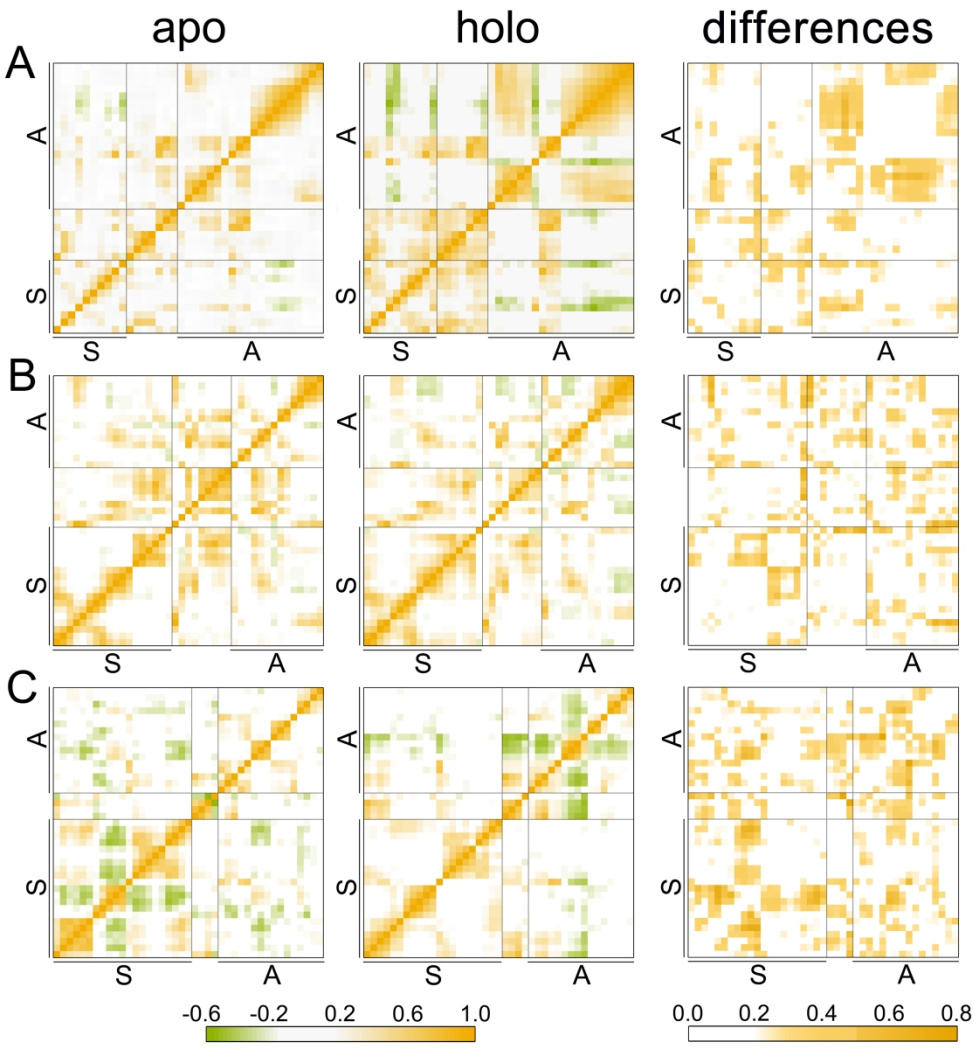


Figure 3. Dynamic cross-correlation maps (DCCM) reveal the extent of correlation for the orthosteric and allosteric sites as well as the shared community between them in the apo and holo forms of TrpRS (A), ERK/MAP kinase (MAPK) (B), and K-Ras4B (C). The absolute values for the differences in DCCM between the apo and holo forms are shown in the right panel. Motion occurring along the same direction is represented by positive correlation (orange), while anti-correlated motion occurring along the opposite direction is represented by negative correlation (green). 'S' and 'A' represent orthosteric and allosteric sites, respectively.

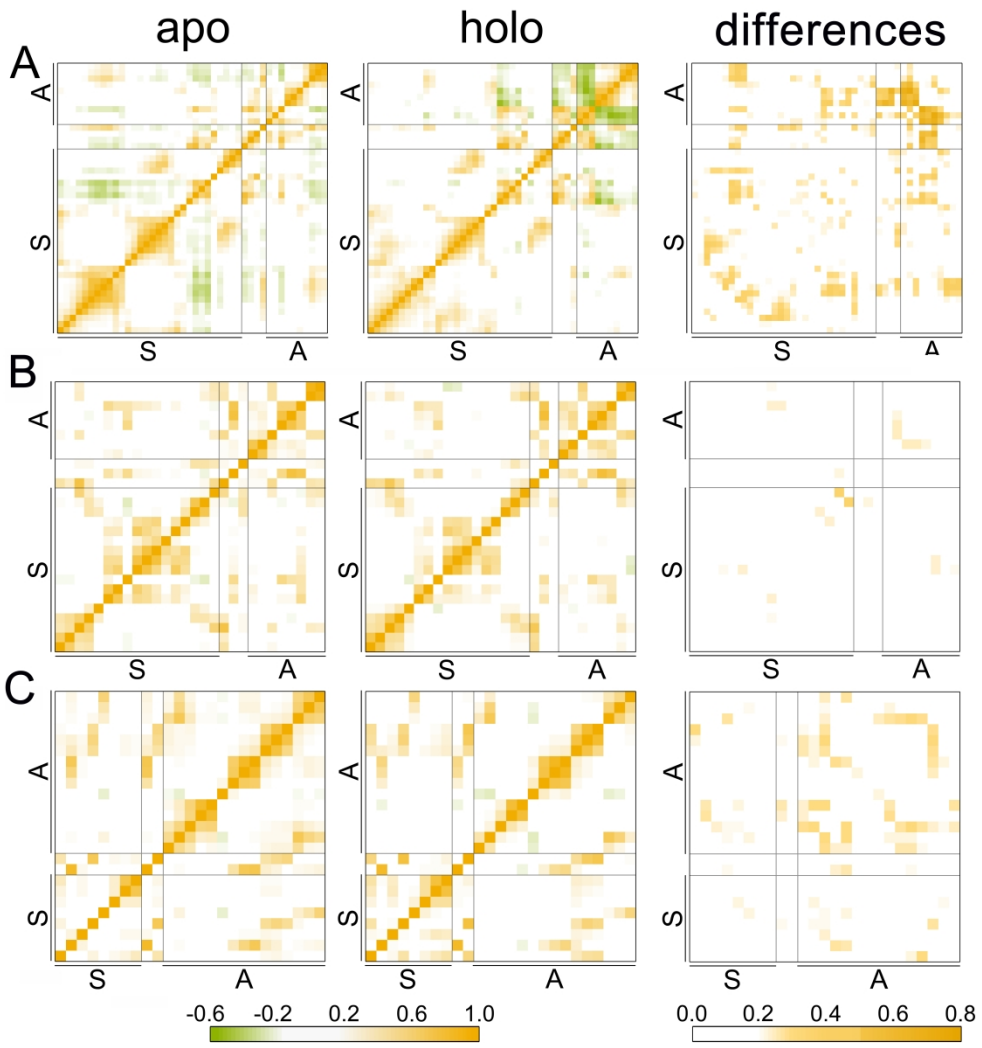


Figure 4. Dynamic cross-correlation maps (DCCM) reveal the extent of correlation for the orthosteric and allosteric sites as well as the shared community between them in the apo and holo forms of GlmU (A), MAO B (B), and MIF (C). The absolute values for the differences in DCCM between the apo and holo forms are shown in the right panel. Motion occurring along the same direction is represented by positive correlation (orange), while anti-correlated motion occurring along the opposite direction is represented by negative correlation (green). 'S' and 'A' represent orthosteric and allosteric sites, respectively.

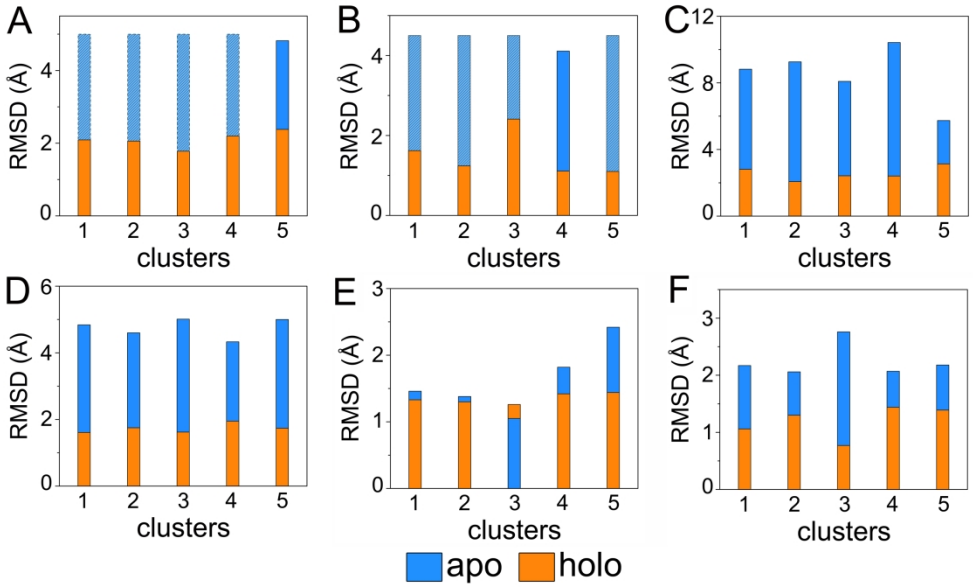


Figure 5. Root-mean-squared-deviation (RMSD) of each allosteric modulator between the native-binding mode and the re-docked binding mode for the five representative structures extracted from the MD trajectory. The blue and orange bars represent the RMSD for the apo and holo forms, respectively. TrpRS (A), ERK/MAP kinase (MAPK) (B), K-Ras4B (C), GlmU (D), MAO B (E), and MIF (F).

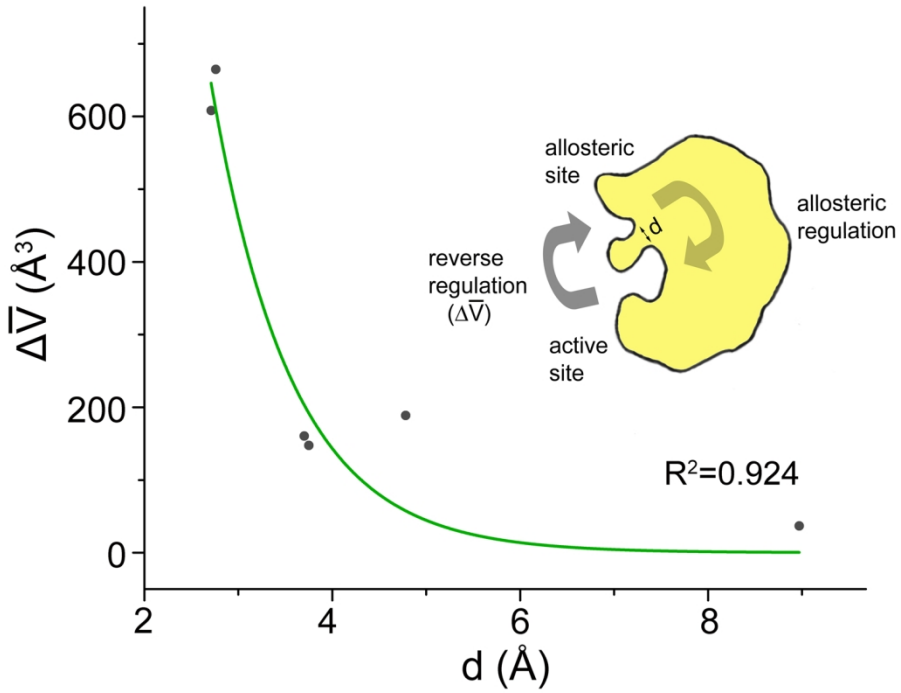


Figure 6. Relationship between the distance of the centroids between allosteric and orthosteric sites and the allosteric site volume.