PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

A comprehensive ensemble model for comparing the allosteric effect of ordered and disordered proteins

Luhao Zhang[1,2], **Maodong Li**[3], **Zhirong Liu**[1,3,4]*

1 College of Chemistry and Molecular Engineering, Peking University, Beijing, China, **2** Department of Chemistry, Princeton University, Princeton, NJ, United States of America, **3** Center for Quantitative Biology, Peking University, Beijing, China, **4** State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Beijing National Laboratory for Molecular Sciences (BNLMS), Peking University, Beijing, China

* LiuZhiRong@pku.edu.cn

OPEN ACCESS

Citation: Zhang L, Li M, Liu Z (2018) A comprehensive ensemble model for comparing the allosteric effect of ordered and disordered proteins. PLoS Comput Biol 14(12): e1006393. https://doi.org/10.1371/journal.pcbi.1006393

Editor: Harel Weinstein, Weill Medical College of Cornell University, UNITED STATES

Received: July 19, 2018

Accepted: November 2, 2018

Published: December 3, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Natural Science Foundation of China (grant 21633001) (URL: http://www.nsfc.gov.cn/english/site_1/index.html) and the Ministry of Science and Technology of China (grant 2015CB910300)(URL: http://www.most.gov.cn/eng/). ZL is the author that received the funding. The funders had no role in study design, data collection and analysis,

Abstract

Intrinsically disordered proteins/regions (IDPs/IDRs) are prevalent in allosteric regulation. It was previously thought that intrinsic disorder is favorable for maximizing the allosteric coupling. Here, we propose a comprehensive ensemble model to compare the roles of both order-order transition and disorder-order transition in allosteric effect. It is revealed that the MWC pathway (order-order transition) has a higher probability than the EAM pathway (disorder-order transition) in allostery, suggesting a complicated role of IDPs/IDRs in regulatory proteins. In addition, an analytic formula for the maximal allosteric coupling response is obtained, which shows that too stable or too unstable state is unfavorable to endow allostery, and is thus helpful for rational design of allosteric drugs.

Author summary

Allosteric effect is an important regulation mechanism in biological processes, where the binding of a ligand at one site of a protein influences the function of a distant site. Conventionally, allostery was thought to originate from structural transition. However, in recent years, intrinsically disordered proteins (IDPs) were found to be widely involved in allosteric regulation in despite of their lack of ordered structure under physiological condition. It is still a mystery why IDPs are prevalent in allosteric proteins and how they differ from ordered proteins in allostery. Here, we propose a comprehensive ensemble model which includes both ordered and disordered states of a two-domain protein, and investigate the role of various state combinations in allosteric effect. By sampling the parameter space, we conclude that disordered proteins are less competitive than ordered proteins in performing allostery from a thermodynamic point of view. The prevalence of IDPs in allosteric regulation is likely determined by all their advantage, but not only by their capacity in endowing allostery.

decision to publish, or preparation of the manuscript.

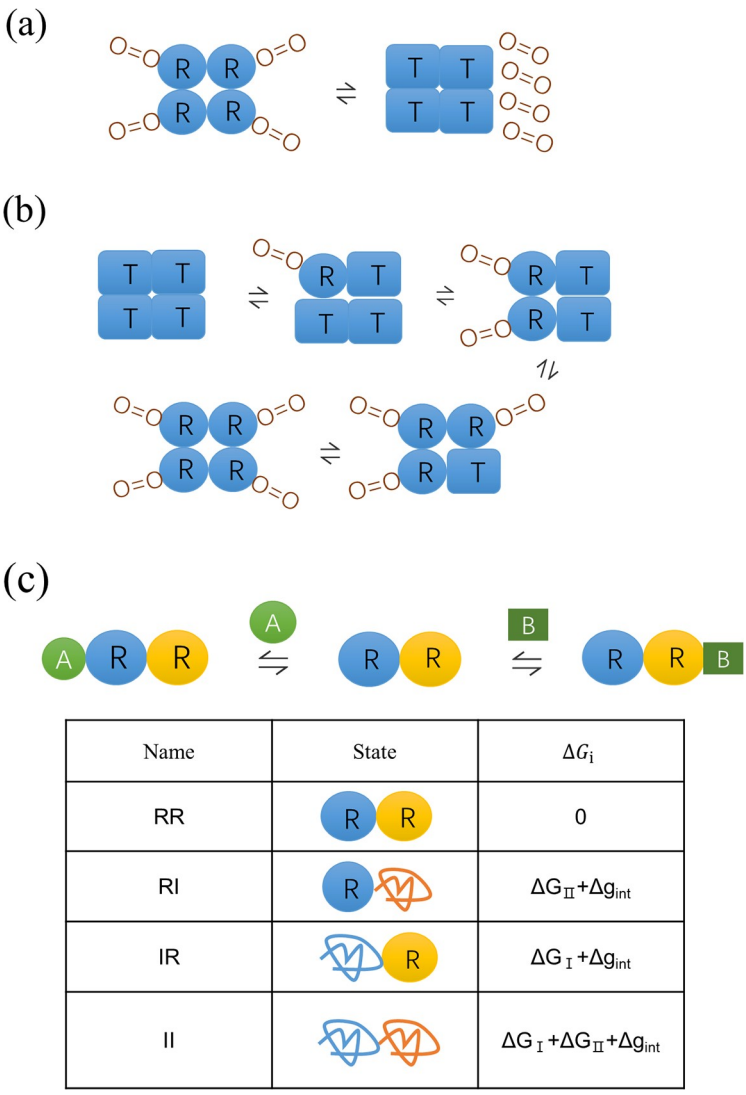
Competing interests: The authors have declared that no competing interests exist.

Introduction

Allosteric regulation is intrinsic to the control of many metabolic and signal-transduction pathways [1]. It is described as the effect that the binding of a ligand at one site of a protein influences the function of a distant site which binds with substrate [2]. In history, several models have been proposed illuminating possible mechanism of allostery. The classical MWC (Monod-Wyman-Changeux) [3] model explained the allosteric effect based on a cooperative conformational transition of protein oligomers. Taking hemoglobin binding with oxygen as an example (Fig 1(A)), the MWC model assumes that four subunits of hemoglobin are simultaneously in either a relaxed state (R state) or a tense state (T state), and oxygens bind preferentially to the R state which shifts the R-T equilibrium. With such a simple assumption, the MWC model nicely explained how the binding of oxygen at one site promotes the binding at a remote site. Later, the KNF (Koshland-Nemethy-Filmer) model [4] has considered finite subunit interactions and proposed a progressive conformational transition of each domain step by step (Fig 1(B)). Both models imply that allosteric processes are closely associated with ligand-driving conformational changes that propagate between the allosterically coupled binding sites. With the development of structural biology, the description of allostery in terms of structure changes was derived [5], and was used to study allosteric proteins such as lactate dehydrogenase [6]. The structure paradigm also leads to the seeking of specific atomic pathway that connects allosteric sites [7–9]. Nevertheless, the discovery of dynamic structure and multiple conformations of proteins, such as multiple orientations of DNA-binding domains of DNA-binding proteins in the absence of DNA [10] and the intermediate conformation of hemoglobin in solution [11], suggests more possibilities beyond the simple two-state models.

The discovery of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) has brought a challenge to the conventional "structure-function" paradigm [13–16]. IDPs/IDRs do not have ordered structures in the free state under physiological conditions, but they are important in biological signaling and regulation [17–25]. IDPs/IDRs possess some advantages over ordered proteins [26], such as high specificity coupled with low affinity useful for reversible signaling interaction [27–30], binding to multiple partners [31,32], and rapid turnover allowing sensitive response to environment changing [13,20,33]. Therefore, they play crucial roles in widespread categories of proteins [23], e.g., scaffold proteins [34], RNA and protein chaperones [35], transcription factors [21], regulation of cellular pathways [36], and the recent liquid-liquid phase separations [37]. In particular, IDPs/IDRs were found to be widely involved in allosteric regulation in despite of their lack of ordered structures [38–48]. Representative examples include enzyme aminoglycoside N-(6')-acetyltransferase II (AAC), which has local unfolding and switching behaviors from positive cooperativity to negative cooperativity upon different temperature [40]; and Doc/Phd toxin-antitoxin system with intrinsic disorder exhibiting complex "conditional cooperativity" character upon different Doc/Phd ratio [41,45].

How can IDPs/IDRs implement allosteric effect under the lack of ordered structures? And why are they so prevalent in allosteric regulation? The answer is related to an emerging new view of allostery based on the general landscape theory of protein structure, where the ligand binding stabilizes specific states and shifts the conformational ensemble [49–51]. The EAM (Ensemble Allostery Model) model used the ensemble view to explain the allostery of IDPs [12,52–55], see Fig 1(C). As an example, it described a two domain system as a four-state ensemble with each domain having ordered (R) and disordered (I) states. The allosteric ligand (A) binds only with the R state of the first domain while the substrate (B) binds only with the R state of the second domain, i.e., the disordered states have no affinity to ligand and substrate. When the interface-interaction free energy between two ordered domains is negative, binding



Name	State	ΔG_i
RR	R R	0
RI	R	$\Delta G_{II}+\Delta g_{int}$
IR	R	$\Delta G_{I}+\Delta g_{int}$
II		$\Delta G_{I}+\Delta G_{II}+\Delta g_{int}$

Fig 1. Schematic diagram of MWC, KNF and EAM models. (a) The MWC model for hemoglobin [3]. All four subunits are simultaneously in either an R (Relaxed) state (with a higher affinity for O_2) or a T (Tense) state (with a lower affinity for O_2). The cooperative effect happens when one O_2 binds to the all-R state, shifts the chemical balance from all-T to all-R, then creates more favorable binding sites for subsequent O_2. (b) The KNF model for hemoglobin [4]. It allows several intermediate states keeping balanced by several chemical equilibrium constants. (c) The EAM model for a two-domain protein [12]. Each domain can be in either an R (Relaxed) state or an I (Disordered) state, resulting in four possible combinations for protein states: RR, RI, IR and II. The corresponding probability P_i of four states is related to their free energy ΔG_i (relative to the RR state) as $P_i = e^{-\Delta G_i/RT}/Q$ where Q is the sum of statistical weights as $Q = \sum_i e^{-\Delta G_i/RT}$. The free energy ΔG_i were listed in the graphics, where ΔG_I and ΔG_{II} are the free energy of unfolding the R state of each domain, and Δg_{int} is the free energy of breaking the interface interactions of ordered complex (RR). One domain (blue) in the R state can bind the allosteric ligand (A) while the other domain (yellow) in the R state can bind the substrate (B). The I state of each domain has no affinity to ligand and substrate.

https://doi.org/10.1371/journal.pcbi.1006393.g001

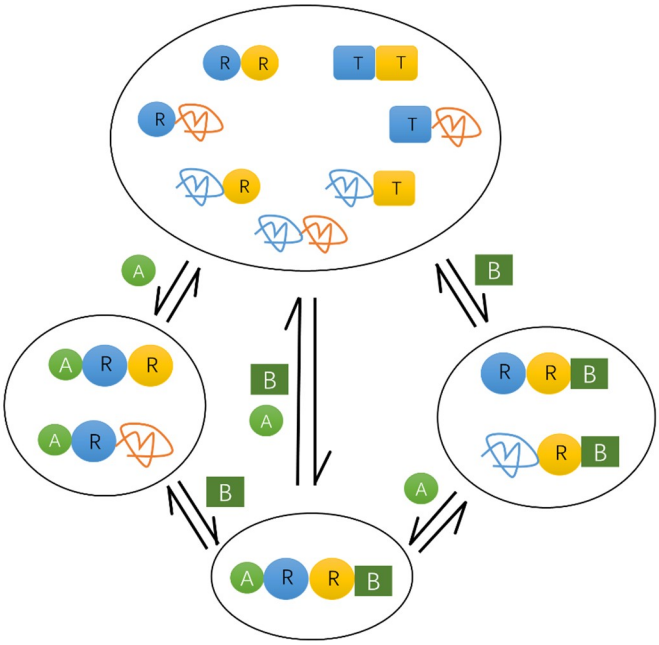
of the ligand A would stabilize the RR state and thus facilitate the binding of the substrate B, resulting in a positive allosteric effect. Similarly, a negative allosteric effect arises when the interface interaction is unfavorable. The EAM model also provided insight in explaining why IDPs/IDRs are so prevalent in allosteric regulation: it was shown that high allosteric intensity is accompanied by high probability of disordered (I) states [12]. However, in investigating the role of protein disorder in allostery, the pristine EAM model considers only the order-disorder

(R-I) transition [12], but lacks the order-order (R-T) transition as that in the MWC model for the allostery of ordered proteins. Therefore, with separate EAM or MWC models, it is impossible to determine whether disordered or ordered proteins are more advantageous in allosteric regulation. To get a full view of competition of ordered and disordered proteins in allosteric effect, here we propose a comprehensive ensemble model considering both order-disorder and order-order transitions. In this comprehensive model, the EAM and MWC mechanisms become two pathways for allostery of the system, and thus their role can be quantitatively evaluated.

Models

The comprehensive ensemble model

Our proposed model describes a two-domain protein system (Fig 2). It compounds the assumptions of both the MWC model and the pristine EAM model. Each domain has three states: R (Relaxed), T (Tense) and I (Disordered). To keep consistent with the MWC model, R and T are assumed to be incompatible with each other and thus the combinations "RT" and "TR" are forbidden in the resulting protein states. Similar to the EAM model, the I state of a domain is disordered and does not have any interface interaction with the adjacent domain, and it does not bind to any ligand or substrate due to the lack of ordered structures. As a result, there are seven possible combinations for protein states, which are listed in Fig 2 with the



Name	State	ΔG_i
RR	R R	0
RI	R	$\Delta G_{R2}+\Delta g_{int,R}$
IR	R	$\Delta G_{R1}+\Delta g_{int,R}$
II		$\Delta G_{R1}+\Delta G_{R2}+\Delta g_{int,R}$
TT	T T	$\Delta G_{RT1}+\Delta G_{RT2}-\Delta g_{int,T}+\Delta g_{int,R}$
TI	T	$\Delta G_{RT1}+\Delta G_{R2}+\Delta g_{int,R}$
IT	T	$\Delta G_{RT2}+\Delta G_{R1}+\Delta g_{int,R}$

Fig 2. Schematic representation of the proposed comprehensive ensemble model, which compounds the assumptions of both the MWC and the EAM models. Each domain (blue and yellow) can be in R (Relaxed), T (Tensed) and I (Disordered) state. R and T are incompatible as assumed in the MWC model, and thus there are now seven possible combinations for protein states. Similar to the EAM model, the disordered I state of a domain does not have any interface interaction with the adjacent domain, and it does not bind to any ligand or substrate due to the lack of ordered structures. The expressions for the free energy (ΔG_i) of each state (relative to RR as the reference state as that in the EAM model) were listed in the absence of ligand and substrate. ΔG_{R1} and ΔG_{R2} are the free energy of unfolding the R state of each domain, and $\Delta g_{int,R}$ and $\Delta g_{int,T}$ are the free energy of breaking the interface interactions in RR and TT, respectively, which were defined in a manner similar to the EAM model. ΔG_{RT1} and ΔG_{RT2} are the free-energy of the R-T transition for each domain. A is allosteric regulation ligand binding to one (blue) domain, and B is the substrate to the other (yellow) domain. A and B are different molecules, i.e., we consider the heterotropic allosteric effect. To enable both positive and negative allosteric effect for ordered proteins, we consider two binding modes for A: it can only bind to the R state of the blue domain (A-R binding mode) (as depicted here), or can only bind to the T state of the blue domain (A-T binding mode). B always binds only to the R state of the yellow domain.

https://doi.org/10.1371/journal.pcbi.1006393.g002

formula of their free energy ΔG_i in the absence of ligand and substrate. Six free energy parameters (ΔG_{R1}, ΔG_{R2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$, ΔG_{RT1}, ΔG_{RT2}) are basic parameters of the model, determining the ensemble distribution. The corresponding probability P_i of each state is related to their free energy ΔG_i as $P_i = e^{-\Delta G_i/RT}/Q$ where Q is the sum of statistical weights as $Q = \sum_i e^{-\Delta G_i/RT}$. The substrate B binds only to the R state of one (yellow) domain. The allosteric ligand A binds to the other (blue) domain but there are two binding modes: in the A-R binding mode A binds only to the R state of the blue domain (as depicted in Fig 2), while it is the A-T binding mode when A binds only to the T state of the blue domain. The two binding modes are taken into account here to enable both positive and negative allosteric effects for ordered proteins (MWC mechanism), making a comparison between the roles of ordered and disordered proteins possible. For example, if we look at a subsystem consisting of RR and TT states, binding of A in the A-R binding mode increases the fraction of the RR state and thus enhances the subsequent binding of B (activation), while that in the A-T binding mode weakens the binding of B (inhibition).

Definitions of contribution of ordered and disordered protein pathways to allostery of the comprehensive ensemble model

Adding allosteric ligand A to the system results in a redistribution of the protein ensemble probabilities, i.e., population shift [56]. The allosteric effect is directly related to probability variation of the states that can bind substrate B due to the adding of A. There were various ways to measure the allosteric response [1,12,57,58]. Following the EAM model [12], here we define the allosteric coupling response (CR) as

$$CR = \frac{P_{X,[A]} - P_{X,[A]=0}}{-\Delta g_{Lig.A}/RT} \tag{1}$$

to quantitatively measure the allosteric intensity for a given system. Here, X denotes the states that can bind B, so $P_{X,[A]}$ is the probability of states that can bind B when there exists ligand A, and $P_{X,[A]=0}$ is the probability when A is absent. The influence of other measurements of allostery will be discussed below. In the comprehensive ensemble model proposed here, for the A-R binding mode we have $P_{X,[A]} = P_{ARR} + P_{RR} + P_{IR}$, and for the A-T binding mode we have $P_{X,[A]} = P_{RR} + P_{IR}$. $\Delta g_{Lig.A}$ is the stabilizing free energy of adding ligand A for the states that can bind A, which is determined as:

$$\Delta g_{Lig.A} = -RT\ln(1 + K_{a,A} \times [A]), \tag{2}$$

where $K_{a,A}$ is the intrinsic equilibrium constant of the binding reaction for A. For example, in the A-R binding mode, $K_{a,A}$ is the association constant for the reactions A + RR = ARR and A + RI = ARI, which gives the equilibrium distributions:

$$\begin{cases} [RR] + [ARR] = [RR] + K_{a,A}[A][RR] = e^{-\Delta g_{Lig.A}/RT}[RR] \\ [RI] + [ARI] = [RI] + K_{a,A}[A][RI] = e^{-\Delta g_{Lig.A}/RT}[RI] \end{cases}, \tag{3}$$

clearly demonstrating the nature of the stabilizing free energy $\Delta g_{Lig,A}$. In our study, we fixed $\Delta g_{Lig.A} = -3.0$ kcal/mol at a physiological temperature of $T = 310.15$ K as in the EAM model unless otherwise specified.

Because the comprehensive model includes all the states of the MWC model and the EAM model, we can also view the comprehensive system consisting of three subsystems: the MWC subsystem, the EAM subsystem and the Others subsystem (Fig 3). If the allostery occurs via the order-order transition within the MWC subsystem (involving the states RR and TT), it was

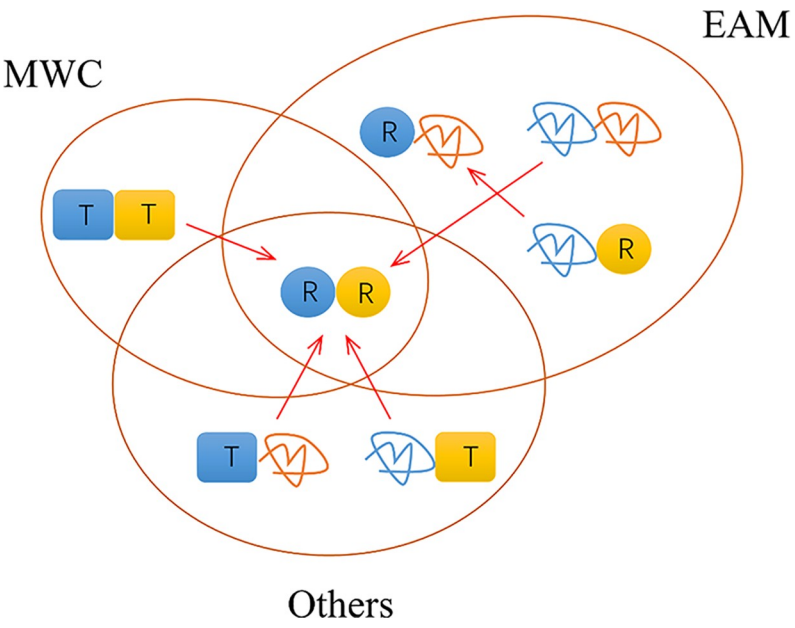


Fig 3. Three allosteric subsystems/pathways in the comprehensive ensemble model. The arrows indicate the population shifts caused by ligand binding which are capable of affecting the substrate binding (see text for details). The A-R binding mode is assumed here. The contribution ratios of pathways to the allostery of the comprehensive system are given in Eqs (4–7).

https://doi.org/10.1371/journal.pcbi.1006393.g003

classified into the MWC pathway. If the allostery occurs via the disorder-order transition within the EAM subsystem (RR, RI, IR and II), it was classified into the EAM pathway. Similar definition applies for the Others pathway involving RR and the remaining states (TI and IT) neglected in the MWC and EAM mechanisms. Binding of the allosteric ligand A in the A-R binding mode causes population shifts of $[\overline{\mathrm{R}}*] \rightarrow [\mathrm{R}*]$, where $\overline{\mathrm{R}}$ represents the non-R domain states (T and I), while * represents any domain states (R, T, I). On the other hand, only the population shifts as $[*\overline{\mathrm{R}}] \rightarrow [*\mathrm{R}]$ could affect the binding of the substrate B. Overall, the population shifts caused by ligand binding which are capable of affecting the substrate binding includes "II→RR", "TT→RR", "TI→RR", "IT→RR" and "IR→RI" (indicated by arrows in Fig 3), which all occurs within the decomposed subsystems here. Therefore, the classification of three pathways is complete and there is no allosteric communication among them. The MWC pathway contains allosteric order-order transition, and the EAM pathway contains allosteric order-disorder transition, while the Others pathway is a mixed one with both order-order and order-disordered transitions, e. g., "TI→RR" is composed of an order-order "T→R" in one domain and an order-disorder "I→R" in the second domain.

The allosteric coupling response (*CR*) of each subsystem can be defined and calculated separately, i.e., to assume each subsystem exists alone. Take the MWC subsystem as an example (under A-R binding mode), we have

$$CR_{\mathrm{MWC}} = \frac{P^{(\mathrm{MWC})}_{\mathrm{RR+ARR},[\mathrm{A}]} - P^{(\mathrm{MWC})}_{\mathrm{RR},[\mathrm{A}]=0}}{-\Delta g_{\mathrm{Lig,A}}/RT}, \tag{4}$$

where the superscript "(MWC)" indicates that the related probabilities of states are defined (normalized) within the MWC subsystem, i.e., $P^{(\mathrm{MWC})}_{\mathrm{RR+ARR},[\mathrm{A}]} = \frac{P_{\mathrm{RR}}+P_{\mathrm{ARR}}}{P_{\mathrm{RR}}+P_{\mathrm{ARR}}+P_{\mathrm{TT}}}$ and $P^{(\mathrm{MWC})}_{\mathrm{RR},[\mathrm{A}]=0} = \frac{P_{\mathrm{RR}}}{P_{\mathrm{RR}}+P_{\mathrm{TT}}}$.

Similarly, *CR* for the EAM subsystem and the Other subsystem are determined by

$$\begin{cases} CR_{\mathrm{EAM}} = \dfrac{P^{(\mathrm{EAM})}_{\mathrm{RR+ARR+IR,[A]}} - P^{(\mathrm{EAM})}_{\mathrm{RR+IR,[A]=0}}}{-\Delta g_{\mathrm{Lig,A}}/RT} \\ CR_{\mathrm{Others}} = \dfrac{P^{(\mathrm{Others})}_{\mathrm{RR+ARR,[A]}} - P^{(\mathrm{Others})}_{\mathrm{RR,[A]=0}}}{-\Delta g_{\mathrm{Lig,A}}/RT} \end{cases}, \tag{5}$$

where $P^{(\mathrm{EAM})}_{\mathrm{RR+ARR+IR,[A]}} = \frac{P_{\mathrm{RR}}+P_{\mathrm{ARR}}+P_{\mathrm{IR}}}{P_{\mathrm{RR}}+P_{\mathrm{ARR}}+P_{\mathrm{RI}}+P_{\mathrm{ARI}}+P_{\mathrm{IR}}+P_{\mathrm{II}}}$ and $P^{(\mathrm{Others})}_{\mathrm{RR+ARR,[A]}} = \frac{P_{\mathrm{RR}}+P_{\mathrm{ARR}}}{P_{\mathrm{RR}}+P_{\mathrm{ARR}}+P_{\mathrm{TI}}+P_{\mathrm{IT}}}$. It is noted that $P^{(\mathrm{MWC})}_{\mathrm{RR+ARR,[A]}} \neq P^{(\mathrm{Others})}_{\mathrm{RR+ARR,[A]}}$ since they are normalized within different subsystems. With a set values of the basic parameters (ΔG_{R1}, ΔG_{R2}, $\Delta g_{\mathrm{int,R}}$, $\Delta g_{\mathrm{int,T}}$, ΔG_{RT1}, ΔG_{RT2}), it is thus straightforward to calculate the probabilities of all the states with and without ligand A, as well as *CR* for the whole system (CR_{tot}) and subsystems (CR_{MWC}, CR_{EAM}, CR_{Others}). The contribution of a pathway to the total allostery of the comprehensive system depends not only on *CR* of the corresponding subsystem, but also on the proportion of the subsystem states in the whole system. Therefore, the contribution ratio of the MWC pathway to the allostery of the comprehensive system is approximately defined as:

$$Weight_{\mathrm{MWC}} = \min(P_{\mathrm{RR+TT,[A]=0}}, P_{\mathrm{RR+ARR+TT,[A]}}) \times CR_{\mathrm{MWC}}/CR_{\mathrm{tot}}. \tag{6}$$

It stands for the weight of the MWC pathway in the allosteric effect. When there are only RR and TT states before adding ligand A, the comprehensive model degenerates to the MWC model and Eq (6) gives $Weight_{\mathrm{MWC}} = 1$. Similarly, for the EAM and the Others pathways, we have:

$$\begin{cases} Weight_{\mathrm{EAM}} = \min(P_{\mathrm{RR+RI+IR+II,[A]=0}}, P_{\mathrm{RR+ARR+RI+ARI+IR+II,[A]}}) \times CR_{\mathrm{EAM}}/CR_{\mathrm{tot}} \\ Weight_{\mathrm{Others}} = \min(P_{\mathrm{RR+TI+IT,[A]=0}}, P_{\mathrm{RR+ARR+TI+IT,[A]}}) \times CR_{\mathrm{Others}}/CR_{\mathrm{tot}} \end{cases}. \tag{7}$$

It is noted that $Weight_{\mathrm{MWC}}$, $Weight_{\mathrm{EAM}}$ and $Weight_{\mathrm{Others}}$ are metrics for three pathways' contributions to allosteric effect of the comprehensive system, but the sum of them is not necessarily equal to 1.0 although the deviation is usually small. Related equations under the A-T binding mode can be found in Supporting Information.

Results

Limits for the maximal allosteric response

With a given set of parameters for protein state stability (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{\mathrm{int,R}}$, $\Delta g_{\mathrm{int,T}}$) and protein-ligand interaction ($\Delta g_{\mathrm{Lig,A}}$) of the proposed comprehensive ensemble model, we can calculate the ensemble distribution, the allosteric coupling response (*CR*) and the contributions of different pathways with the formulism described above. *CR* as a function of $\Delta g_{\mathrm{int,R}}$ and $\Delta g_{\mathrm{int,T}}$ is shown in Fig 4A and 4B as a case example when the other parameters are fixed. It reveals that combination of $\Delta g_{\mathrm{int,R}}$ and $\Delta g_{\mathrm{int,T}}$ is required to maximize the allosteric effect. Under the A-R binding mode, the model can afford both positive ($CR > 0$) and negative ($CR < 0$) allosteric effects, while there is only negative effect under the A-T binding mode. The achieved highest *CR* is about 0.17. To have a global inspection on the occurring probability of allostery, we assume the stability free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{\mathrm{int,R}}$, $\Delta g_{\mathrm{int,T}}$) vary randomly between [−8, +8] kcal/mol, and determine the distribution of *CR* for two binding modes with $\Delta g_{\mathrm{Lig,A}} = -3$ cal/mol (Fig 4C and 4D). For the majority of parameter sets, the resulting allostery is weak, giving a sharp peak at $CR = 0$ for both binging modes (Fig 4C and 4D). Actually, only 6.3% of parameter sets produce $|CR| > 0.1$ under the A-R binding mode. Remarkably, *CR* has the boundaries at around ±0.172. In other words, no matter how the state stabilities of protein are optimized, it is impossible to achieve a *CR* value higher than 0.172.

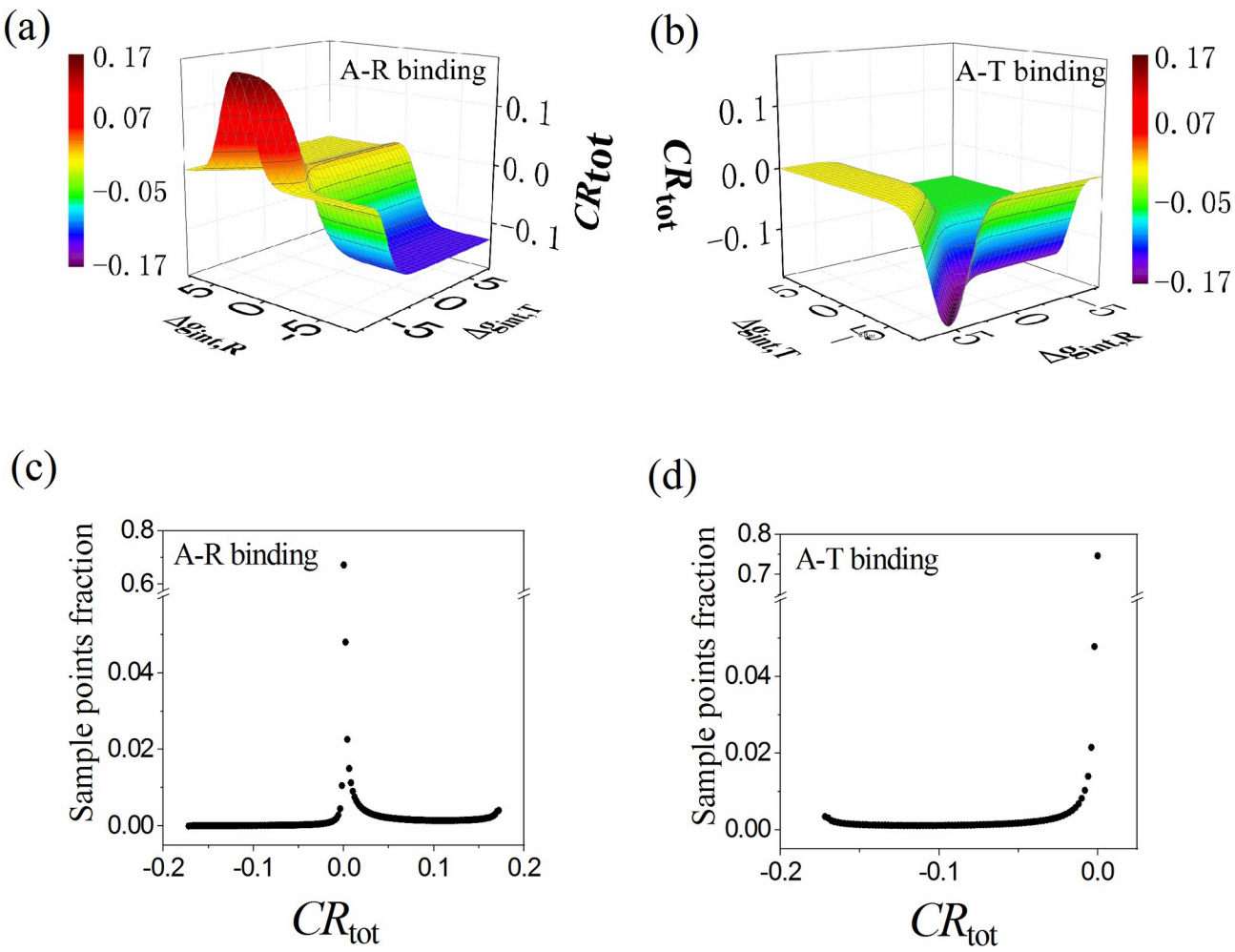


Fig 4. The allosteric coupling response (*CR*) of the comprehensive ensemble model. (a,b) *CR* as a function of $\Delta g_{\mathrm{int,R}}$ and $\Delta g_{\mathrm{int,T}}$ (in kcal/mol) when the other parameters are fixed (chosen to produce notable allostery) as: $\Delta G_{\mathrm{R1}} = -1.0$, $\Delta G_{\mathrm{R2}} = 1.3$, $\Delta G_{\mathrm{RT1}} = 1.0$, $\Delta G_{\mathrm{RT2}} = 3.0$ (all in units of kcal/mol). Note that for A-T binding mode there is no activated allosteric effect. (c,d) Distribution of *CR* when the stability free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{\mathrm{int,R}}$, $\Delta g_{\mathrm{int,T}}$) vary randomly with an equal probability density between −8 and +8 kcal/mol. The A-R binding mode is adopted in (a,c) and the A-T binding mode is adopted in (b,d) with $\Delta g_{\mathrm{Lig,A}} = -3$ cal/mol.

https://doi.org/10.1371/journal.pcbi.1006393.g004

The boundary limits of *CR* can be well explained in an analytic way. Take the MWC model as a simplified example, there are two states (RR and TT state) with only one stability parameter ($\Delta G_i \equiv G_{\mathrm{RR}} - G_{\mathrm{TT}}$), which determines the probability of RR state without ligand to be:

$$P_{\mathrm{RR}} = \frac{e^{-\Delta G_i/RT}}{e^{-\Delta G_i/RT}+1} = 1 - P_{\mathrm{TT}}. \tag{8}$$

CR can then be written as a function of P_{RR} and $\Delta g_{\mathrm{Lig,A}}$ as

$$\begin{aligned} CR &= \left[\frac{P_{\mathrm{RR}} e^{-\Delta g_{\mathrm{Lig,A}}/RT}}{P_{\mathrm{TT}} + P_{\mathrm{RR}} e^{-\Delta g_{\mathrm{Lig,A}}/RT}} - P_{\mathrm{RR}}\right] \cdot \frac{1}{-\Delta g_{\mathrm{Lig,A}}/RT} \\ &= \left[\frac{P_{\mathrm{RR}} e^{-\Delta g_{\mathrm{Lig,A}}/RT}}{1 - P_{\mathrm{RR}} + P_{\mathrm{RR}} e^{-\Delta g_{\mathrm{Lig,A}}/RT}} - P_{\mathrm{RR}}\right] \cdot \frac{1}{-\Delta g_{\mathrm{Lig,A}}/RT} \end{aligned} \tag{9}$$

under the A-R binding mode. The relations among P_{RR}, ΔG_{i} and *CR* are plotted in Fig 5(A) for

$\Delta g_{\mathrm{Lig,A}} = -3$ kcal/mol. *CR* is equal to 0 at either $P_{\mathrm{RR}} = 0$ or $P_{\mathrm{RR}} = 1$, i.e., too stable and too unstable RR state are unfavorable to allostery. *CR* reaches its maximum of about 0.172 at $P_{\mathrm{RR}} = 0.081$. P_{RR} depends on ΔG_i in a switch-like manner. A great many ΔG_i values give P_{RR} close to 0 or 1, and result in small *CR* and weak allostery. This provide a clue in understanding the dominant peak at $CR = 0$ in Fig 4C and 4D. Based on Eq (9), the maximization of *CR* can be solved analytically with $\frac{\partial CR}{\partial P_{\mathrm{RR}}} = 0$ to give

$$CR_{\max} = \frac{\left(e^{-\Delta g_{\mathrm{Lig,A}}/2RT} - 1\right)^2}{(-\Delta g_{\mathrm{Lig,A}}/RT)\left(e^{-\Delta g_{\mathrm{Lig,A}}/RT} - 1\right)} \tag{10}$$

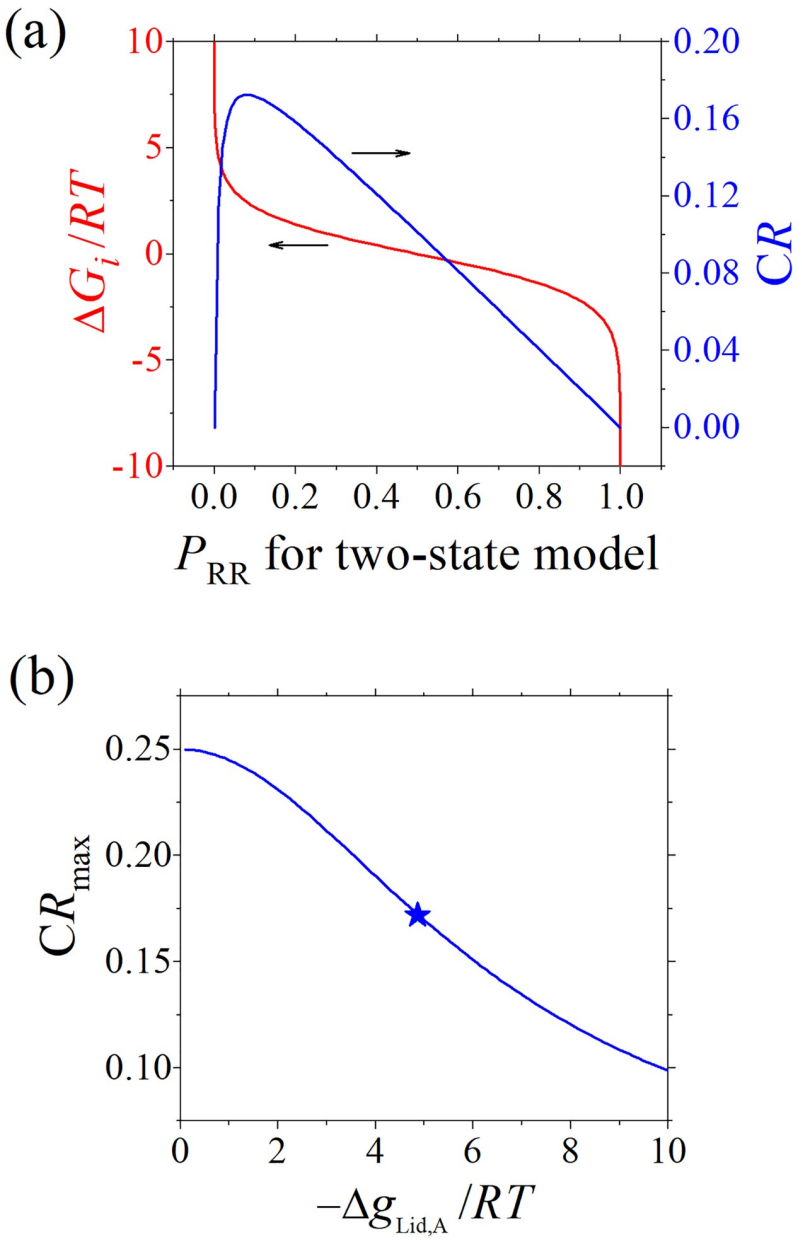


Fig 5. Limits of CR. (a) The relations among P_{RR}, ΔG_i and *CR* for the MWC model with two states, plotted according to Eqs 8 and 9. (b) The *CR* maximum (via optimizing state stabilities of proteins) as a function of $-\Delta g_{\mathrm{Lig,A}}/RT$, plotted according to Eq (10) which is valid for the comprehensive ensemble model as well as the MWC and EAM models. The star represents the data point for $\Delta g_{\mathrm{Lig,A}} = -3$ kcal/mol and $T = 310.15$ K used in other figures.

https://doi.org/10.1371/journal.pcbi.1006393.g005

at the optimized P_{RR} as

$$P_{RR}^{(opt.)} = \frac{e^{-\Delta g_{Lig,A}/2RT} - 1}{e^{-\Delta g_{Lig,A}/RT} - 1}. \quad (11)$$

Eq (10) keeps valid for the comprehensive ensemble model (see Supporting Information). CR_{max} is plotted in Fig 5(B) as a function of $-\Delta g_{Lig,A}/RT$ (note that $\Delta g_{Lig,A} < 0$). It decreases with increasing $-\Delta g_{Lig,A}/RT$, and reaches a value of 0.172 at $\Delta g_{Lig,A} = -3$ kcal/mol and $T = 310.15$ K, being consistent with the observation in Fig 4. Eq (10) gives an analytical result for the limits of CR when the state stabilities of protein are optimized, and would be useful in studying the allosteric capacity of proteins.

The weight of MWC pathway is significantly higher than that of EAM pathway

The weights of three pathways (MWC, EAM and Others) in the allostery of the comprehensive system are numerically analyzed when the stability free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$) vary randomly between [–8, +8] kcal/mol. The resulting average weights are shown in Fig 6(A) as functions of CR. For positive allosteric effect ($CR > 0$), the weight of the MWC pathway is much larger than the EAM one, indicating the MWC pathway holds an advantage over the EAM pathway in this case. For negative allosteric effect, CR under the A-R binding mode mainly comes from the EAM pathway, while under the A-T binding mode CR mainly comes from the MWC and Others pathways. The reason is that when A binds with R, in the MWC subsystem the decrease of RR state is not allowed and thus its weight is almost zero or even negative based on Eq (6), while an IR→RI transition of EAM pathway dominates the negative allosteric response. On the other hand, when A binds with T, it has no effect in the state distribution in the EAM subsystem thus its weight is always zero.

The capacity of the MWC or the EAM pathway for allostery depends on not only their weights in a comprehensive system (Fig 6(A)) but also the possibility of the system to afford an allosteric effect ($P(CR)$, see Fig 4(B)). Therefore, the possibility for allosteric effect with CR undertaken by the MWC pathway can be calculated as

$$P_{MWC}(CR) = \overline{Weight_{MWC}}(CR) \times P(CR). \quad (12)$$

It describes the probability of a randomly chosen parameter set to possess an allosteric effect CR via the MWC pathway. Formula for the EAM and Others pathways can be similarly written. The calculated results are shown in Fig 6(B). $P_{MWC}(CR)$ and $P_{Others}(CR)$ has sharp peaks near the positive allostery limit CR_{max} in the A-R binding mode and near the negative allostery limit $-CR_{max}$ in the A-T binding mode, which will be discussed in detail below. More importantly, if we take a simplified approach by adding curves in the A-R and A-T binding modes for each pathway, $P_{MWC}(CR)$ is much larger than $P_{EAM}(CR)$ for strong allosteric effects. Therefore, the MWC pathway is more important in allosteric effects than the EAM pathway based on the comprehensive ensemble model.

Probability of strong allostery first increases and then decreases when the ΔG_i range increases

The distribution of allostery and pathway contribution were investigated above when the free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$) of the comprehensive model vary randomly in a range of [–8, +8] kcal/mol. The results may change under a different range. In Fig 7(A), the possibilities for an allosteric effect to occur with CR undertaken by three pathways are plotted under various variation range [$-\Delta G_{max}$, $+\Delta G_{max}$] of the free-energy

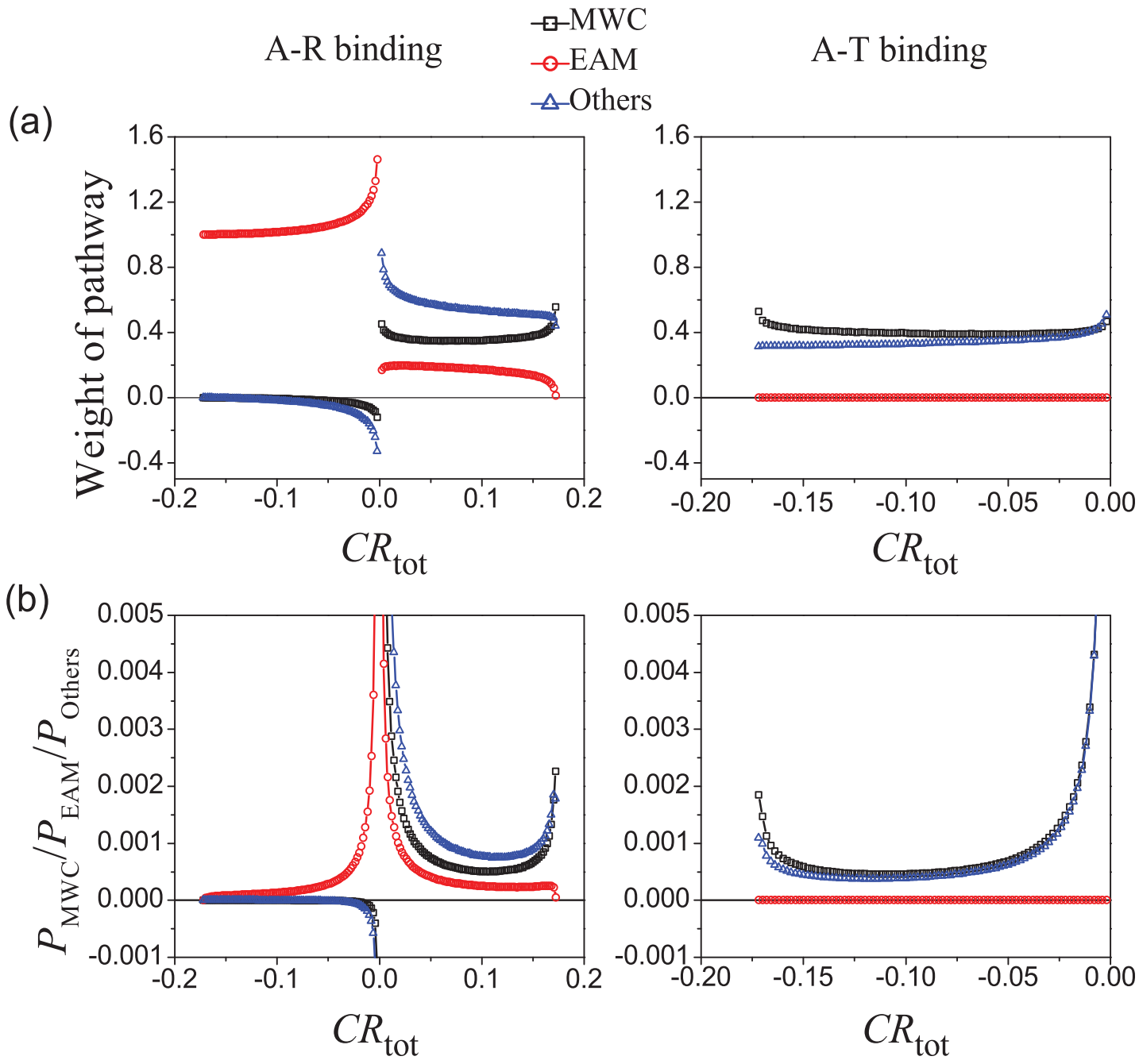


Fig 6. Contributions of three pathways (MWC, EAM, Others) in the comprehensive ensemble model. The stability free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$) vary randomly with an equal probability density between −8 and +8 kcal/mol, resulting in different systems (samples). (a) The average weights of pathways as functions of CR_{tot}. The pathway weights of a system (sample) are calculated based on Eqs 6 and 7. (b) The capacity of three pathways for allostery, being calculated with Eq (12). The A-R binding mode is adopted in left panels and the A-T binding mode is adopted in right panels with $\Delta g_{Lig,A} = -3$ kcal/mol. It is noted that a large portion of samples practically have $CR = 0$ and the pathway contributions are ill-defined with Eqs 6 and 7, which are thus ignored.

https://doi.org/10.1371/journal.pcbi.1006393.g006

parameters. The sharp peaks of P_{MWC} and P_{Others} near the positive allostery limit (CR_{max} = 0.172) observed previously are absent when the variation range (ΔG_{max}) is small, e.g., ΔG_{max} = 1 kcal/mol. In Fig 7(B), the probabilities of $CR > 0.171$ for three pathways are plotted as a function of ΔG_{max}. It clearly shows that the MWC and the Others pathways have a similar

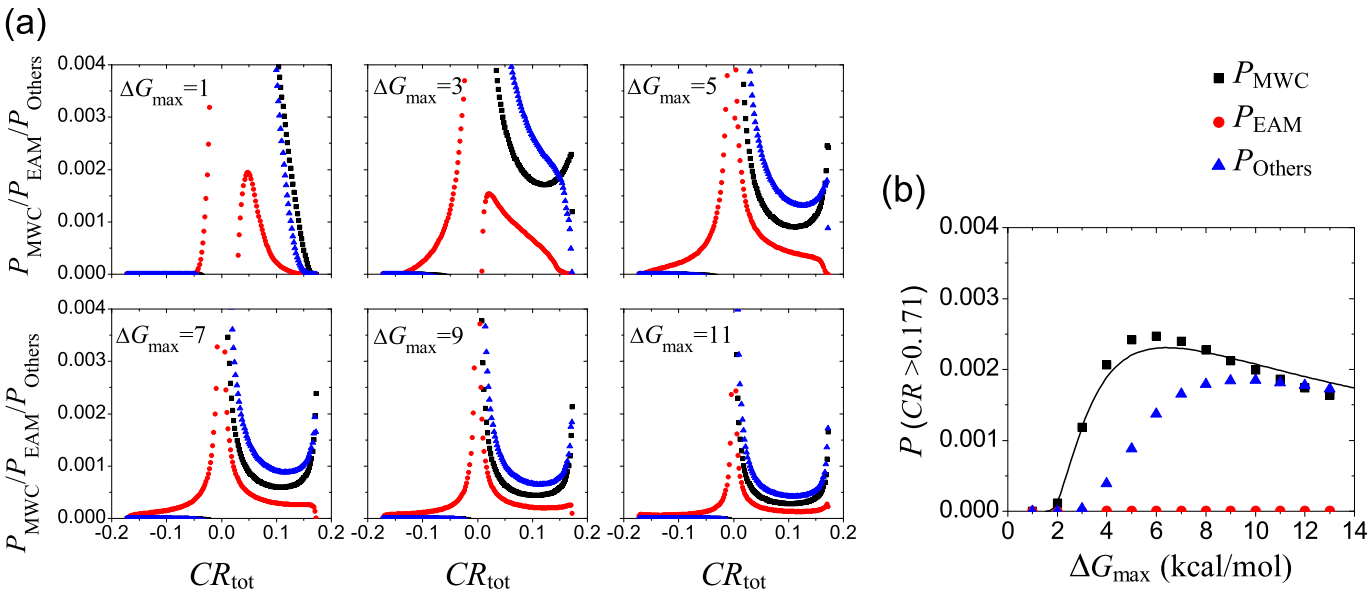


Fig 7. The influence of the variation range (ΔG_{max}, in a unit of kcal/mol) for free-energy parameters (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$) of the comprehensive ensemble model with the A-R binding mode. (a) The possibilities for three pathways [calculated with Eq (12)] obtained at different [$-\Delta G_{max}$, $+\Delta G_{max}$] range. (b) The probabilities of $CR > 0.171$ as a function of ΔG_{max}.

https://doi.org/10.1371/journal.pcbi.1006393.g007

tendency: it first equals to zero before a critical ΔG_{max} (which is smaller for the MWC pathway), then increases quickly, and finally decreases slowly.

The feature observed in Fig 7 can be qualitatively explained based on the simplified two-state model (Fig 5). The maximal *CR* is achieved at $P_{RR} = 0.081$, which corresponds to a free energy difference of $\Delta G_i (\equiv G_{RR} - G_{TT}) = 1.6$ kcal/mol. When the variation range of the free-energy parameters is small, the resulting ΔG_i cannot reach the optimized value for the maximal *CR*, giving the zero value in Fig 7(B) and the absence of the sharp peak near CR_{max} in the panel with $\Delta G_{max} = 1$ kcal/mol in Fig 7(A). When the variation range of the free-energy parameters is large enough, although the optimized value of ΔG_i can be always satisfied at some values of parameter sets, the total number of possible values increases with the variation range, and thus the probability of maximal *CR*, defined as the ratio between the number of optimized parameter value sets to that of the total number, would decreases with increasing the variation range as observed in Fig 7(B).

Two-state transition is the main mechanism for strong allostery

The comprehensive ensemble model includes seven states and three subsystems/pathways. How do they coordinate in fulfilling the allosteric effect? For example, do the pathways repeal each other in a system? How many states play significant role in a system? Here, we investigate the interplay between different states and different subsystems/pathways in the allosteric process.

To measure the mixing extend of subsystems and pathways, we classify each system case (with a certain set of ΔG_i values) into one of four categories: single subsystem with single pathway (S,S), single subsystem with mixing pathways (S,M), mixing subsystems with single pathway (M,S), and mixing subsystems with mixing pathways (M,M). If the sum of state probability for any subsystem is larger than 0.99 before and after adding ligand, it is classified into single subsystem; otherwise it belongs to mixing subsystem. Single pathway is defined for the case where the weight of one pathway is larger than 0.99 and the absolute value of weights for other pathways are less than 0.01; otherwise it belongs to mixing pathway. For example, if a

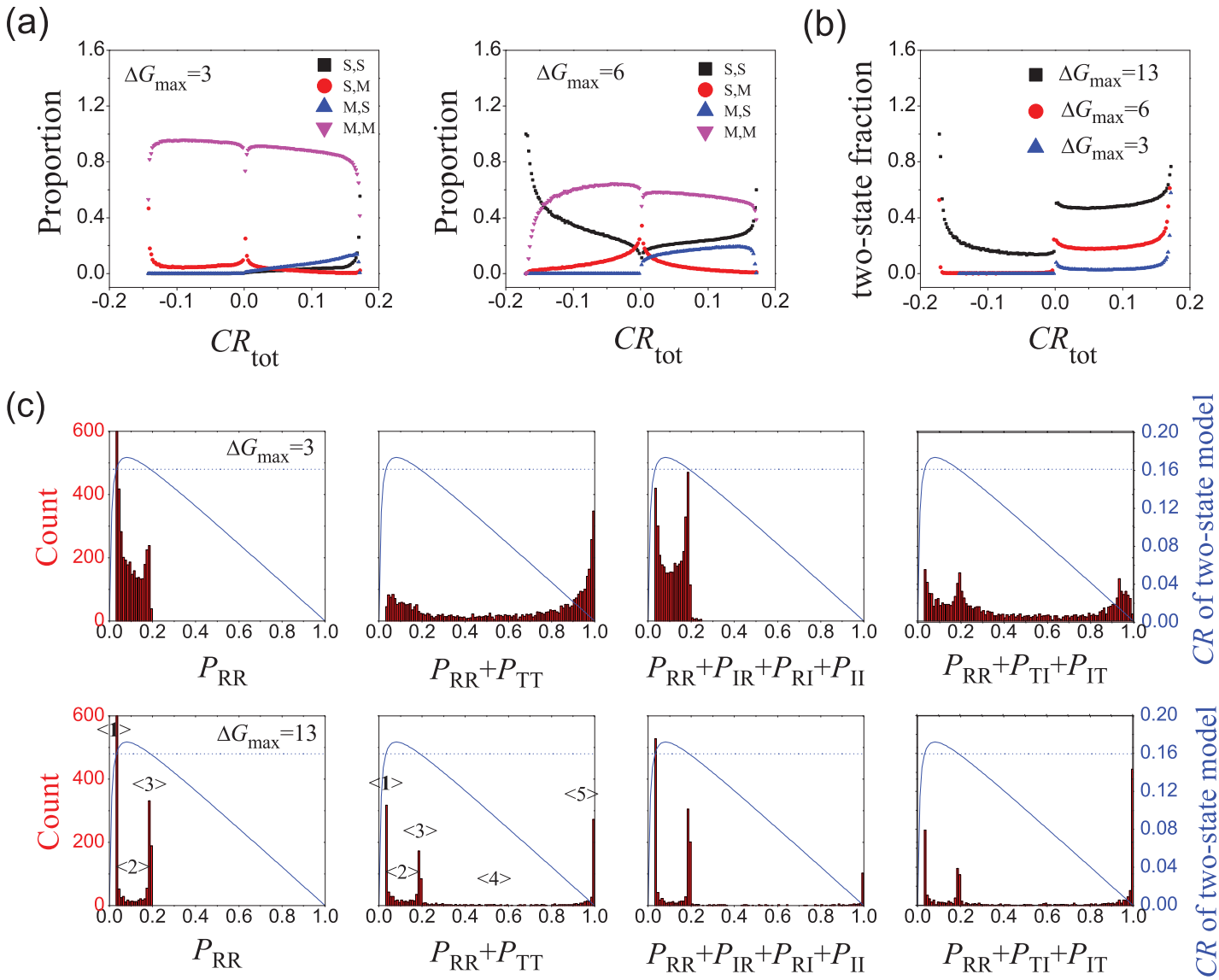


Fig 8. Interplay between different states and subsystems/pathways in the comprehensive ensemble model. The A-R binding mode is adopted. (a) Proportion of four categories [single subsystem with single pathways (S,S), single subsystem with mixing pathways (S,M), mixing subsystems with single pathways (M,S), and mixing subsystems with mixing pathways (M,M)] in systems with *CR*. (b) The proportion of systems with two-state transition. (c) Distribution of P_{RR} and state probability of three subsystems ($P_{RR} + P_{TT}$ for the MWC pathway, $P_{RR} + P_{IR} + P_{RI} + P_{II}$ for the EAM pathway, and $P_{RR} + P_{TI} + P_{IT}$ for the Other pathway) for systems with $CR \approx 0.16$. ΔG_{max} = 3 or 13 kcal/mol. The theoretical $CR \sim P_{RR}$ curve (blue line) for the two-state model is also plotted using Eq (9). The horizontal dashed line represents $CR = 0.16$.

https://doi.org/10.1371/journal.pcbi.1006393.g008

system only contain RR and TT states, then it simply belongs to the (S,S) category. The results are shown in Fig 8(A). When the variation range (ΔG_{max}) of free-energy parameters is small, mixing subsystems with mixing pathways (M,M) dominate in most cases. But when ΔG_{max} is larger, the proportion of single subsystem with single pathway (S,S) increases while the (M,M) type decreases. More importantly, the (S,S) proportion increases with increasing $|CR|$. The system tends to behave as pure subsystem with pure pathway mechanism at strong allostery.

A clearer angle of view is to look at the proportion of systems that implement allostery via a simple mechanism of two-state transition. Here we specify a system to have two-state

transition mechanism if the probability sum of two certain states of the given system is larger than 0.99 both before and after binding with ligand A. Possible two-state transition for positive allosteric effect includes "II→RR", "TT→RR", "TI→RR" and "IT→RR". For negative allosteric effect, the only possible two-state transition is "IR→RI". The proportion of systems with simple two-state transition is shown in Fig 8(B). With larger ΔG_{max}, the proportion of two-state transition is higher. The proportion has a sharp peak at $\pm CR_{max}$. Therefore, two-state transition is the major mechanism for strong allosteric even in the comprehensive ensemble model.

The existence of two-state transition and single subsystem/pathway are also reflected in the state distribution patterns. The distributions of RR and states of three subsystems are shown in Fig 8(C) for systems with $CR \approx 0.16$. The distribution of P_{RR} has two obvious peaks labeled with <**1**> and <**3**>. In Fig 8(C) we also plot the theoretical $CR \sim P_{RR}$ curve for the two-state model for convenience's sake. The crossing points between the $CR \sim P_{RR}$ curve and the horizontal line of $CR = 0.16$ give the P_{RR} values to achieve an allosteric effect of $CR = 0.16$ in the two-state model. The obtained P_{RR} values of the crossing points coincide with the peak position at <**1**> and <**3**> of the simulated P_{RR} distribution, suggesting that the strong allostery (with $CR = 0.16$) of the comprehensive model mainly occurs in a two-state model mechanism (note that RR exists in all possible two-state transition for positive CR including "II→RR", "TT→RR","TI→RR" and "IT→RR"). There is also some nonzero P_{RR} distribution (<**2**>) between two peaks, which is expected to have CR higher than 0.17 in the two-state model. The reason for that is the introducing of additional IR and RI population would decrease CR (see Supplementary Material). It also explains the intriguing result that there is no distribution outside <**1**>&<**3**>, for P_{RR} outside cannot give CR as big as 0.16. When ΔG_{max} increases to 13 kcal/mol, P_{RR} distribution enriches at <**1**>&<**3**> and reduces at <**2**>, suggesting an enrichment of two-state transition mechanism. Similarly, for the distribution of the MWC pathway states, the $P_{RR} + P_{TT}$ peaks at <**1**>&<**3**> correspond to the systems dominated by other pathways (EAM or Others) so that $P_{RR} + P_{TT} = P_{RR}$ and the peak positions are identical to that for P_{RR}. At <**5**>, $P_{RR} + P_{TT} = 1$ corresponds to the systems dominated by the MWC pathway. <**2**> and <**4**> mean hybridized cases. Results for the population distribution of the EAM and Others subsystems are similar. They confirm that strong allostery in the comprehensive ensemble mode is dominated by single pathway and the two-state transition mechanism.

The conclusions are similar under other measures of allostery and interaction schemes

In addition to the allosteric coupling response (CR) we adopted above, there were various ways to measure the allosteric response [1,57,58]. For example, a widely-used one is the thermodynamic allosteric efficacy α defined as [57]

$$\alpha = \frac{K_{\text{bound}}}{K_{\text{unbound}}}, \tag{13}$$

where K_{bound} and K_{unbound} are the equilibrium constants for the binding with substrate B when the protein is bound or unbound to the allosteric ligand A, respectively. For the A-R binding mode of the current comprehensive ensemble model, α can be calculated as

$$\alpha = \frac{P_{RR}}{P_{R*}} \Big/ \frac{P_{*R}}{P_{**}}, \tag{14}$$

where * indicates the summation over all possible domain states, e.g., $P_{R^*} = P_{RR} + P_{RI}$. Different from CR which counts the probability difference and thus arbitrarily penalizes the allosteric stabilization of very low probability states (for example, a $P = 0.0001$ state becoming a $P = 0.01$

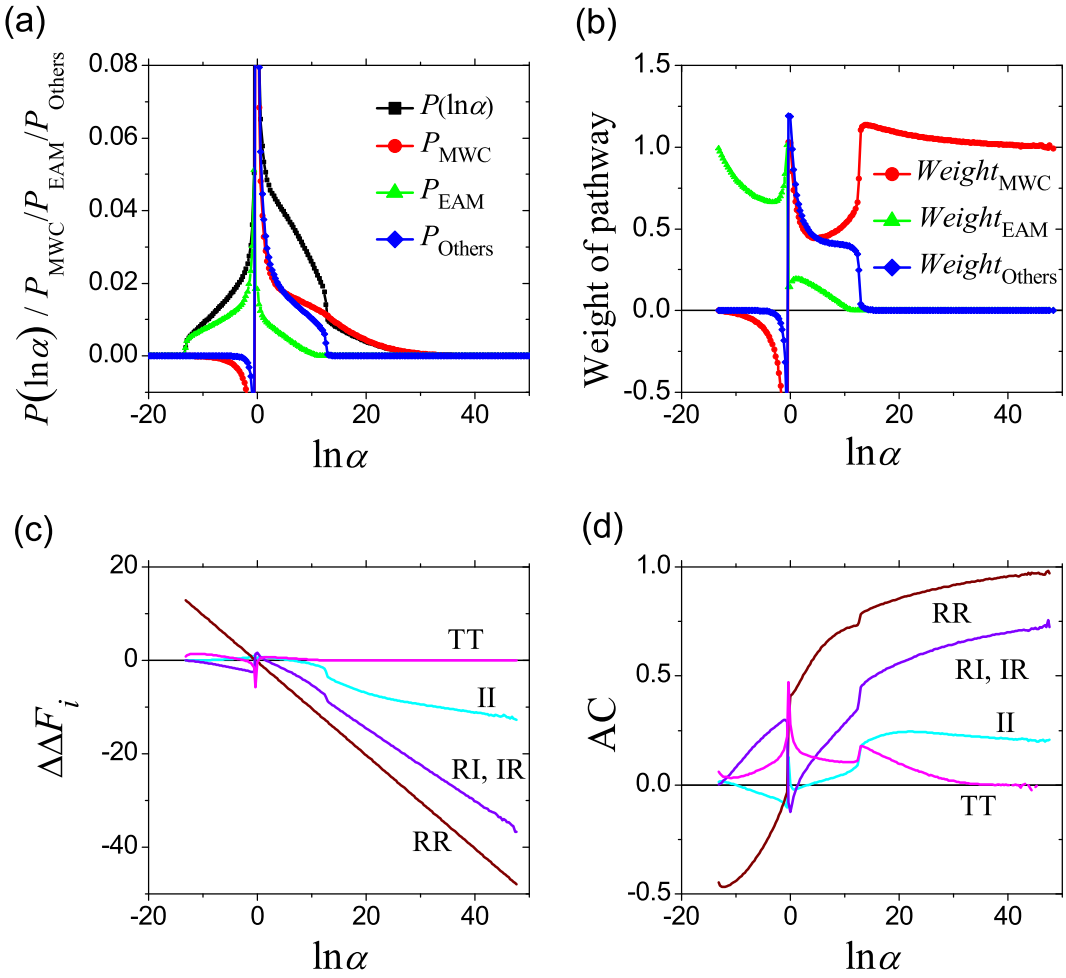


Fig 9. Analysis results with other measures of allostery. (a) Distribution of lnα where α is the thermodynamic allosteric efficacy defined in Eq (13), as well as the capacity of three pathways. (b) The average weights of pathways as functions of lnα. (c) The average thermodynamic coupling function [$\Delta\Delta F_i$, as defined in Eq (15), and in a unit of RT] at some states as functions of lnα. (d) The average normalized allosteric coupling [AC, as defined in Eq (16)] of some states as functions of lnα. The variation range (ΔG_{max}) for free-energy parameters is 8 kcal/mol.

https://doi.org/10.1371/journal.pcbi.1006393.g009

state would be considered "weak"), α counts the fold-change in probabilities and is directly related to thermodynamic energy cycle of allostery [57]. To clarify whether the choice of allosteric measure would affect the conclusions, we examine the distribution of α and the contributions from three pathways in Fig 9A and 9B. For positive allosteric effect ($\ln\alpha > 0$), the distribution probability decreases with increasing α. More importantly, strong allostery is dominantly contributed by the MWC pathway. In other words, the advantage of MWC pathway is even more prominent when α is used instead of CR to measure allostery.

The pristine EAM model mainly describes the order-disorder transition [12]. It was later extended to become a general two-state model of allostery applicable for any system exhibiting conformational change (e.g., IDPs, folded proteins, or combinations thereof) [59,60]. The comprehensive ensemble model proposed in this study can be regarded as a further generalization to the three-state case where we combined the insight of the pristine EAM model on order-disorder transition (the lack of interface interaction for I) and that of the MWC model on order-order transition (the incompatible interface between R and T). Actually, the latter point is vital for the achievement of allostery. This can be well demonstrated by the

thermodynamic coupling function introduced by LeVine *et al.* [58,61]:

$$\Delta\Delta F(x,y) = -RT\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{15}$$

in describing the allostery landscape, where x and y represents the state of the allosteric and the functional sites, respectively, i.e., the state of the first domain and the second domain. For a simplified case where all free energy parameters are set to 0, we have $P_{\mathrm{RR}} = P_{\mathrm{IR}} = P_{\mathrm{RR}} = P_{\mathrm{II}} = P_{\mathrm{TT}} = P_{\mathrm{TI}} = P_{\mathrm{IT}} = 1/7$, $P_{\mathrm{R}^*} = P_{^*\mathrm{R}} = P_{\mathrm{T}^*} = P_{^*\mathrm{T}} = 2/7$ and $P_{\mathrm{I}^*} = P_{^*\mathrm{I}} = 3/7$, so it yields that (in unit of RT) $\Delta\Delta F_{\mathrm{RR}} = \Delta\Delta F_{\mathrm{TT}} = -\ln\left[\frac{1}{7}/\left(\frac{2}{7}\right)^2\right] \approx -0.56$, $\Delta\Delta F_{\mathrm{RI}} = \Delta\Delta F_{\mathrm{IR}} = -\ln\left[\frac{1}{7}/\left(\frac{2}{7}\cdot\frac{3}{7}\right)\right] \approx -0.15$ and $\Delta\Delta F_{\mathrm{II}} = -\ln\left[\frac{1}{7}/\left(\frac{3}{7}\right)^2\right] \approx 0.25$. This may explain why the model seems to utilize the order-order pathway more often than the order-disorder one–it is a natural result of fully excluding the RT and TR states. Actually, under the A-R binding mode, α defined in Eq (14) is related to $\Delta\Delta F$ in Eq (15) as $\ln\alpha = -\frac{\Delta\Delta F_{\mathrm{RR}}}{RT}$. A numerical result in Fig 9(C) confirms such a prediction. $\Delta\Delta F$ has a natural normalization when x and y are discrete, and a normalized allosteric coupling was proposed by LeVine *et al.* as [58]:

$$AC(x,y) = -\frac{\Delta\Delta F(x,y)}{\Delta\Delta F_{\max}(x,y)} = \frac{\ln(p(x)p(y))}{\ln(p(x,y))} - 1, \tag{16}$$

which ranges between 1 and −1 and its magnitude describes what fraction of the maximal allostery is contributing to the free energy of the joint state. Analysis on AC is given in Fig 9(D). AC of RR is close to 1, indicating its central role in allosteric response.

The choice of the form of interaction is also important and may affect the results. The pristine EAM model utilizes a simplifying assumption that the state I does not involve any interface interaction. On the other hand, other schemes are possible. For example, LeVine and Weinstein have proposed an Allosteric Ising Model (AIM) where an Ising-like interaction was adopted to greatly facilitate analyses [57]. Here, we incorporate AIM to demonstrate the influence of the interaction scheme. In the spirit of AIM, the allosteric ligand A can be regarded as regular component (in a role similar to the protein domains) of the system and adopts two states: on and off. Therefore, if we incorporate the essence of AIM into the MWC model, the system has four possible states: (on/off)(R)(R), (on/off)(T)(T), while the incorporation into the EAM results in six states: (on/off)(R)(R/I), (off)(I)(R/I). With an Ising-like interaction among components, α is given as

$$\begin{aligned} \alpha_{\mathrm{MWC}} &= e^{-\frac{4E_{AR}}{RT}} \\ \alpha_{\mathrm{EAM}} &= e^{-\frac{2E_{RR}}{k_{\mathrm{B}}T}} \frac{e^{\frac{E_{AR}+E_{RR}}{RT}} + e^{-\frac{E_{AR}+E_{RR}}{RT}}}{e^{\frac{E_{AR}-E_{RR}}{RT}} + e^{-\frac{E_{AR}-E_{RR}}{RT}}} \end{aligned} \tag{17}$$

where E_{AR} and E_{RR} represent the Ising interaction parameters of the ligand-domain and domain-domain interface. The distributions are shown in Fig 10(A) when E_{AR} and E_{RR} randomly vary in a range of $[\Delta G_{\max}, -\Delta G_{\max}]$ with $\Delta G_{\max} = 3\ RT$. The MWC model has a much higher probability to achieve strong allostery than the EAM model even if the Ising-like interaction of AIM was incorporated. For the comprehensive ensemble model with three states for each domain, a Potts model (the generalized of the Ising model to > 2 states per component) can be introduced to describe the domain-domain interaction [the ligand has only two states

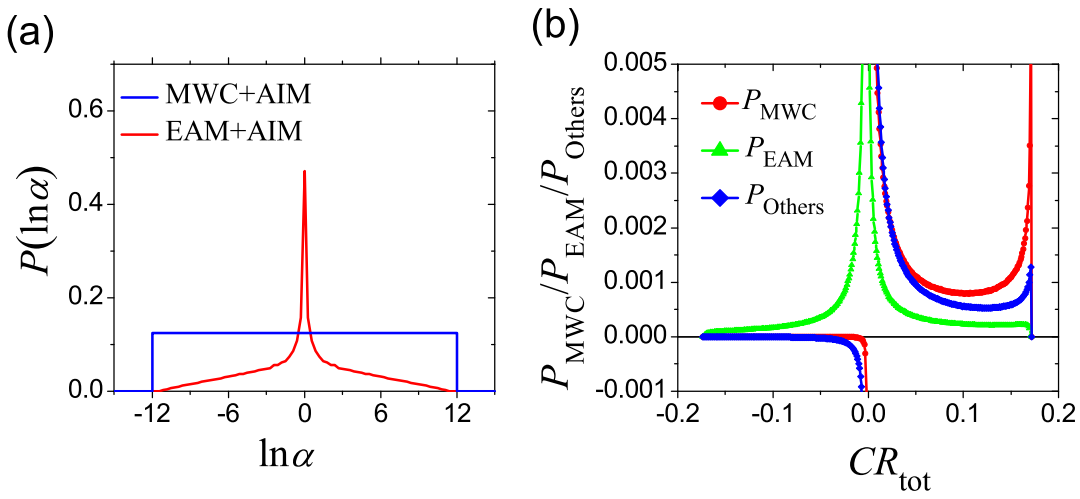


Fig 10. The properties under Ising-like interface interactions. (a) Distribution of lnα when the Ising-like interface interactions of the Allosteric Ising Model (AIM) were introduced to combine with the MWC and EAM models. The variation range of Ising interface energy is ΔG_{max} = 3 RT. (b) The capacity of three pathways in the comprehensive ensemble model when the interface interactions are Ising-like. The variation range (ΔG_{max}) for free-energy parameters is 8 kcal/mol.

https://doi.org/10.1371/journal.pcbi.1006393.g010

and it is assumed to be unaffected by the Potts model here, i.e., Eq (3) is reserved], and the results are shown in Fig 10(B). The peaks near the positive allostery limit CR_{max} are higher than those before introducing Ising-like interaction (comparing Fig 10(B) and Fig 6(B)), indicating that an Ising-like interaction will enhance the probability of strong allostery. What is unaffected is the trend that $P_{MWC}(CR)$ is much larger than $P_{EAM}(CR)$ for strong allosteric effects. In other words, the importance of the MWC pathway over the EAM pathway is not influenced by the interaction scheme in this case.

Discussion

Possible reasons for the prevalence of IDPs/IDRs in allosteric regulation

IDPs/IDRs appear in much higher amounts in regulatory proteins [21,25]. They are also widely involved in allosteric processes [38–43,45,47], although it is unclear whether they are involved more widely than order proteins do. A possible explain for the prevalence of IDPs/IDRs in allosteric regulation was provided by the EAM model which suggested that intrinsic disorder can maximize the ability to allosteric coupling [12]. However, our comprehensive ensemble model reveals that the order-disorder transition (EAM mechanism) is actually less competitive than the order-order transition (MWC mechanism) in affording allosteric effects, especially the strong allostery. It shows that the reasons for the prevalence of IDPs/IDRs in allosteric regulation are more complicated than previously thought in EAM. Our work does not give a complete answer for it, but we provide some discussion and comments here.

Firstly, in our study we assumed that the free energy parameters of conformation change and domain-domain interaction (ΔG_{R1}, ΔG_{R2}, ΔG_{RT1}, ΔG_{RT2}, $\Delta g_{int,R}$, $\Delta g_{int,T}$) vary randomly with an equal probability density between [$-\Delta G_{max}$, $+\Delta G_{max}$]. In real proteins it does not have to be like this. The difficulty (probability) to modify order-order and order-disorder transitions is likely different. Specifically, to tune the protein stability difference between two similar order structures (R and T in the MWC model) via mutation would be more difficult than to tune the stability difference between order and disordered structures (R and I in the EAM model), because in the latter case this can be accomplished via breaking or strengthening a

residue-residue interaction that is present in ordered structure but absent in disordered structure. Therefore, a possible reason for the prevalence of IDPs/IDRs in allosteric regulation is their convenience in modifying state stability.

Secondly, IDPs/IDRs possess various advantages over ordered proteins [26,62], such as saving genome resources via multi-binding pattern or creating large binding surface, overcoming steric effect in binding, accelerating binding speed, achieving high specificity with low affinity, and facilitating posttranslational modifications. The prevalence of IDPs/IDRs in allosteric regulation is determined by all their advantage, but not only by their capacity in endowing allostery. Work combining experimental data and bioinformatics analyses would be helpful to compare ordered and disordered proteins' importance in allosteric regulation.

Thirdly, due to the lack of systematic survey, at present it is still unclear whether IDPs/IDRs are more or less prevalent in allostery than ordered proteins. They might play a complementary role to expand the proteome of allostery even if IDPs/IDRs do not exceed ordered proteins.

Lastly, allosteric effects with maximal *CR* may be not the pursuing goal. Allostery with different strength would have different applications. For example, allosteric effect that are not too strong is beneficial in ensuring safer dosing [63].

Conclusions

In this work, we proposed a comprehensive ensemble model to study the role of order-order and order-disorder transitions in allosteric effect. An analytic equation for the maximal allosteric coupling response (*CR*) was derived, which shows that too stable or too unstable state is unfavorable to achieve allostery. By sampling the parameter space, it was revealed that the order-order transition (MWC) mechanism has a higher possibility in allostery than the order-disorder transition (EAM) mechanism, and the remaining mixed (Others) mechanism involving both order-order and order-disorder transitions also has a high possibility close to the MWC one. In addition, two-state transition is the primary mechanism when allostery is strong although there are seven states in the model. The work not only provided insight in understand the prevalence of IDPs/IDRs in allosteric regulation, but is also helpful for rational design of allosteric drugs.

Supporting information

S1 Text. *CR* formula of subsystems under A-T binding mode and the general proof on the maximal *CR* in the comprehensive ensemble model.
(PDF)

Acknowledgments

The authors thank Huaiqing Cao, Miao Yu and Hao Ruan for helpful discussions.

Author Contributions

Conceptualization: Zhirong Liu.

Data curation: Luhao Zhang.

Formal analysis: Luhao Zhang, Zhirong Liu.

Funding acquisition: Zhirong Liu.

Investigation: Luhao Zhang, Maodong Li.

Methodology: Luhao Zhang.

Project administration: Luhao Zhang, Zhirong Liu.

Resources: Maodong Li.

Supervision: Zhirong Liu.

Validation: Luhao Zhang, Zhirong Liu.

Visualization: Luhao Zhang, Zhirong Liu.

Writing – original draft: Luhao Zhang.

Writing – review & editing: Zhirong Liu.

References

1. Fenton AW (2008) Allostery: an illustrated definition for the 'second secret of life'. Trends In Biochemical Sciences 33: 420–425. https://doi.org/10.1016/j.tibs.2008.05.009 PMID: 18706817
2. Monod J, Changeux JP, Jacob F (1963) Allosteric proteins and cellular control systems. Journal Of Molecular Biology 6: 306–329. PMID: 13936070
3. Monod J, Wyman J, Changeux JP (1965) On nature of allosteric transitions—a plausible model. Journal Of Molecular Biology 12: 88–118. PMID: 14343300
4. Koshland DE, Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5: 365–385. PMID: 5938952
5. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood? Nature Structural Biology 6: 351–358. https://doi.org/10.1038/7586 PMID: 10201404
6. Iwata S, Kamata K, Yoshida S, Minowa T, Ohta T (1994) T-state and R-state in the crystals of bacterial l-lactate dehydrogenase reveal the mechanism for allosteric control. Nature Structural Biology 1: 176–185. PMID: 7656036
7. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286: 295–299. PMID: 10514373
8. Tang S, Liao J-C, Dunn AR, Altman RB, Spudich JA, et al. (2007) Predicting allosteric communication in myosin via a pathway of conserved residues. Journal Of Molecular Biology 373: 1361–1373. https://doi.org/10.1016/j.jmb.2007.08.059 PMID: 17900617
9. Yu M, Ma X, Cao H, Chong B, Lai L, et al. (2018) Singular value decomposition for the correlation of atomic fluctuations with arbitrary angle. Proteins-Structure Function and Bioinformatics: 86: 1075–1087.
10. Bochkareva E, Belegu V, Korolev S, Bochkarev A (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. EMBO Journal 20: 612–618. https://doi.org/10.1093/emboj/20.3.612 PMID: 11157767
11. Lukin JA, Kontaxis G, Simplaceanu V, Yuan Y, Bax A, et al. (2003) Quaternary structure of hemoglobin in solution. Proceedings of the National Academy of Sciences of the United States of America 100: 517–520. https://doi.org/10.1073/pnas.232715799 PMID: 12525687
12. Hilser VJ, Thompson EB (2007) 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: 8311–8315. https://doi.org/10.1073/pnas.0700329104 PMID: 17494761
13. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. Journal of Molecular Biology 293: 321–331. https://doi.org/10.1006/jmbi.1999.3110 PMID: 10550212
14. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, et al. (2001) Intrinsically disordered protein. Journal of Molecular Graphics & Modelling 19: 26–59.
15. Huang Y-Q, Liu Z-R (2010) Intrinsically disordered proteins: the new sequence-structure-function relations. Acta Physico-Chimica Sinica 26: 2061–2072.
16. Tompa P (2011) Unstructural biology coming of age. Current Opinion In Structural Biology 21: 419–425. https://doi.org/10.1016/j.sbi.2011.03.012 PMID: 21514142
17. Huang Y, Yoon M-K, Otieno S, Lelli M, Kriwacki RW (2015) The activity and stability of the intrinsically disordered Cip/Kip protein family are regulated by non-receptor tyrosine kinases. Journal Of Molecular Biology 427: 371–386. https://doi.org/10.1016/j.jmb.2014.11.011 PMID: 25463440

18. Wallin S (2017) Intrinsically disordered proteins: structural and functional dynamics. Research and Reports in Biology 8: 7–16.
19. Csizmok V, Follis AV, Kriwacki RW, Forman-Kay JD (2016) Dynamic protein interaction networks and new structural paradigms in signaling. Chemical Reviews 116: 6424–6462. https://doi.org/10.1021/acs.chemrev.5b00548 PMID: 26922996
20. Uversky VN (2002) Natively unfolded proteins: A point where biology waits for physics. Protein Science 11: 739–756. https://doi.org/10.1110/ps.4210102 PMID: 11910019
21. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, et al. (2006) Intrinsic disorder in transcription factors. Biochemistry 45: 6873–6888. https://doi.org/10.1021/bi0602718 PMID: 16734424
22. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. Journal Of Molecular Recognition 18: 343–384. https://doi.org/10.1002/jmr.747 PMID: 16094605
23. Uversky VN (2013) A decade and a half of protein intrinsic disorder: Biology still waits for physics. Protein Science 22: 693–724. https://doi.org/10.1002/pro.2261 PMID: 23553817
24. Li M, Sun T, Jin F, Yu D, Liu Z (2016) Dimension conversion and scaling of disordered protein chains. Molecular Biosystems 12: 2932–2940. https://doi.org/10.1039/c6mb00415f PMID: 27440558
25. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. Journal of Proteome Research 6: 1882–1898. https://doi.org/10.1021/pr060392u PMID: 17391014
26. Liu Z, Huang Y (2014) Advantages of proteins being disordered. Protein Science 23: 539–550. https://doi.org/10.1002/pro.2443 PMID: 24532081
27. Huang Y, Liu Z (2013) Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chemistry-a European Journal 19: 4462–4467.
28. Zhou H-X (2012) Intrinsic disorder: signaling via highly specific but short-lived association. Trends In Biochemical Sciences 37: 43–48. https://doi.org/10.1016/j.tibs.2011.11.002 PMID: 22154231
29. Wong ETC, Na D, Gsponer J (2013) On the importance of polar interactions for complexes containing intrinsically disordered proteins. Plos Computational Biology 9: e1003192. https://doi.org/10.1371/journal.pcbi.1003192 PMID: 23990768
30. Bhattacherjee A, Wallin S (2013) Exploring protein-peptide binding specificity through computational peptide screening. Plos Computational Biology 9: e1003277. https://doi.org/10.1371/journal.pcbi.1003277 PMID: 24204228
31. Karush F (1950) Heterogeneity of the binding sites of bovine serum albumin. Journal Of The American Chemical Society 72: 2705–2713.
32. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of $p21^{(Waf1/Cip1/Sdi1)}$ in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. Proceedings of the National Academy of Sciences of the United States of America 93: 11504–11509. PMID: 8876165
33. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, et al. (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. Proceedings of the National Academy of Sciences of the United States of America 104: 9650–9655. https://doi.org/10.1073/pnas.0702580104 PMID: 17522259
34. Cortese MS, Uversky VN, Dunker AK (2008) Intrinsic disorder in scaffold proteins: Getting more from less. Progress In Biophysics & Molecular Biology 98: 85–106.
35. Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. FASEB Journal 18: 1169–1175. https://doi.org/10.1096/fj.04-1584rev PMID: 15284216
36. Xue B, Dunker AK, Uversky VN (2012) The roles of intrinsic disorder in orchestrating the Wnt-pathway. Journal Of Biomolecular Structure & Dynamics 29: 843–861.
37. Li P, Banjade S, Cheng H-C, Kim S, Chen B, et al. (2012) Phase transitions in the assembly of multivalent signalling proteins. Nature 483: 336–340. https://doi.org/10.1038/nature10879 PMID: 22398450
38. Tompa P (2014) Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. Chemical Reviews 114: 6715–6732. https://doi.org/10.1021/cr4005082 PMID: 24533462
39. Reichheld SE, Yu Z, Davidson AR (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. Proceedings of the National Academy of Sciences of the United States of America 106: 22263–22268. https://doi.org/10.1073/pnas.0911566106 PMID: 20080791
40. Freiburger LA, Baettig OM, Sprules T, Berghuis AM, Auclair K, et al. (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. Nature Structural & Molecular Biology 18: 288–294.

41. Garcia-Pino A, Balasubramanian S, Wyns L, Gazit E, De Greve H, et al. (2010) Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell 142: 101–111. https://doi.org/10.1016/j.cell.2010.05.039 PMID: 20603017

42. Smock RG, Gierasch LM (2009) Sending signals dynamically. Science 324: 198–203. https://doi.org/10.1126/science.1169377 PMID: 19359576

43. Krishnan N, Koveal D, Miller DH, Xue B, Akshinthala SD, et al. (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. Nature Chemical Biology 10: 558–566. https://doi.org/10.1038/nchembio.1528 PMID: 24845231

44. Liu J, Nussinov R (2008) Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design. Proceedings of the National Academy of Sciences of the United States of America 105: 901–906. https://doi.org/10.1073/pnas.0707401105 PMID: 18195360

45. Garcia-Pino A, De Gieter S, Talavera A, De Greve H, Efremov RG, et al. (2016) An intrinsically disordered entropic switch determines allostery in Phd-Doc regulation. Nature Chemical Biology 12: 490–496. https://doi.org/10.1038/nchembio.2078 PMID: 27159580

46. Li M, Cao H, Lai L, Liu Z (2018) Disordered linkers in mutli-domain allosteric proteins: entropic effect to favor the open state or enhanced local concentration to favor the closed state? Protein Science 27: 1600–1610. https://doi.org/10.1002/pro.3475 PMID: 30019371

47. Berlow RB, Dyson HJ, Wright PE (2018) Expanding the paradigm: intrinsically disordered proteins and allosteric regulation. Journal Of Molecular Biology 430: 2309–2320. https://doi.org/10.1016/j.jmb.2018.04.003 PMID: 29634920

48. Li M, Liu Z (2016) Dimensions, energetics, and denaturant effects of the protein unstructured state. Protein Science 25: 734–747. https://doi.org/10.1002/pro.2865 PMID: 26683260

49. Kern D, Zuiderweg ERP (2003) The role of dynamics in allosteric regulation. Current Opinion In Structural Biology 13: 748–757. PMID: 14675554

50. Gunasekaran K, Ma BY, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins-Structure Function and Bioinformatics 57: 433–443.

51. Tsai C-J, Nussinov R (2014) A unified view of "how allostery works". Plos Computational Biology 10: e1003394. https://doi.org/10.1371/journal.pcbi.1003394 PMID: 24516370

52. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508: 331–339. https://doi.org/10.1038/nature13001 PMID: 24740064

53. Choi JH, Laurent AH, Hilser VJ, Ostermeier M (2015) Design of protein switches based on an ensemble model of allostery. Nature Communications 6: 1–9.

54. Motlagh HN, Li J, Thompson EB, Hilser VJ (2012) Interplay between allostery and intrinsic disorder in an ensemble. Biochemical Society Transactions 40: 975–980. https://doi.org/10.1042/BST20120163 PMID: 22988850

55. Maillard RA, Liu T, Beasley DWC, Barrett ADT, Hilser VJ, et al. (2014) Thermodynamic mechanism for the evasion of antibody neutralization in flaviviruses. Journal Of The American Chemical Society 136: 10315–10324. https://doi.org/10.1021/ja503318x PMID: 24950171

56. Liu J, Nussinov R (2016) Allostery: an overview of its history, concepts, methods, and applications. Plos Computational Biology 12: e1004966. https://doi.org/10.1371/journal.pcbi.1004966 PMID: 27253437

57. LeVine MV, Weinstein H (2015) AIM for allostery: using the Ising model to understand information processing and transmission in allosteric biomolecular systems. Entropy 17: 2895–2918. https://doi.org/10.3390/e17052895 PMID: 26594108

58. Cuendet MA, Weinstein H, LeVine MV (2016) The allostery landscape: quantifying thermodynamic couplings in biomolecular systems. Journal of Chemical Theory and Computation 12: 5758–5767. https://doi.org/10.1021/acs.jctc.6b00841 PMID: 27766843

59. Hilser VJ, Wrabl JO, Motlagh HN (2012) Structural and energetic basis of allostery. In: Rees DC, editor. Annual Review of Biophysics, Vol 41. pp. 585–609. https://doi.org/10.1146/annurev-biophys-050511-102319 PMID: 22577828

60. Motlagh HN, Hilser VJ (2012) Agonism/antagonism switching in allosteric ensembles. Proceedings of the National Academy of Sciences of the United States of America 109: 4134–4139. https://doi.org/10.1073/pnas.1120519109 PMID: 22388747

61. LeVine MV, Cuendet MA, Razavi AM, Khelashvili G, Weinstein H (2018) Thermodynamic coupling function analysis of allosteric mechanisms in the human dopamine transporter. Biophysical Journal 114: 10–14. https://doi.org/10.1016/j.bpj.2017.10.030 PMID: 29153319

62. Huang Y, Liu Z (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: A critical assessment of the "fly-casting" mechanism. Journal Of Molecular Biology 393: 1143–1159. https://doi.org/10.1016/j.jmb.2009.09.010 PMID: 19747922

63. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, et al. (2016) Emerging computational methods for the rational discovery of allosteric drugs. Chemical Reviews 116: 6370–6390. https://doi.org/10.1021/acs.chemrev.5b00631 PMID: 27074285