

Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein

Yong Wang¹, Chun Tang², Erkang Wang^{1*}, Jin Wang^{1,3,4*}

1 State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, China, **2** State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan, Hubei, China, **3** College of Physics, Jilin University, Changchun, Jilin, China, **4** Department of Chemistry, Physics and Applied Mathematics, State University of New York at Stony Brook, Stony Brook, New York, United States of America

Abstract

An increasing number of biological machines have been revealed to have more than two macroscopic states. Quantifying the underlying multiple-basin functional landscape is essential for understanding their functions. However, the present models seem to be insufficient to describe such multiple-state systems. To meet this challenge, we have developed a coarse grained triple-basin structure-based model with implicit ligand. Based on our model, the constructed functional landscape is sufficiently sampled by the brute-force molecular dynamics simulation. We explored maltose-binding protein (MBP) which undergoes large-scale domain motion between open, apo-closed (partially closed) and holo-closed (fully closed) states responding to ligand binding. We revealed an underlying mechanism whereby major induced fit and minor population shift pathways co-exist by quantitative flux analysis. We found that the hinge regions play an important role in the functional dynamics as well as that increases in its flexibility promote population shifts. This finding provides a theoretical explanation of the mechanistic discrepancies in MBP protein family. We also found a functional “backtracking” behavior that favors conformational change. We further explored the underlying folding landscape in response to ligand binding. Consistent with earlier experimental findings, the presence of ligand increases the cooperativity and stability of MBP. This work provides the first study to explore the folding dynamics and functional dynamics under the same theoretical framework using our triple-basin functional model.

Citation: Wang Y, Tang C, Wang E, Wang J (2012) Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein. *PLoS Comput Biol* 8(4): e1002471. doi:10.1371/journal.pcbi.1002471

Editor: Ray Luo, University of California, Irvine, United States of America

Received: December 16, 2011; **Accepted:** February 26, 2012; **Published:** April 19, 2012

Copyright: © 2012 Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JW acknowledges support from the National Science Foundation (grant no. 21190040 and 11174105) and EW are supported by the National Natural Science Foundation of China (Grants 90713022 and 20735002) and by the Ministry of Science and Technology of China (2009CB930100 and 2010CB933600), and CT is supported by the Chinese Academy of Sciences (KJCX2-EW-W05). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: jin.wang.1@stonybrook.edu (JW); ekwang@ciac.jl.cn (EW)

Introduction

Biomolecular function is executed through conformational dynamics at physiological conditions, ranging from small fluctuations in atomic positions to large movements of parts of or even entire molecules [1,2]. It has been recognized that large-scale domain rearrangement is involved in processes like protein folding, molecular recognition, enzyme catalysis, signal transduction, transcriptional regulation and allostery, and plays an important role in biomolecular machines, such as the ribosome, transporter, molecular chaperones, enzyme, and molecular motors, carrying out their respective functions [3].

Though local conformational changes can be detected using experimental probes such as NMR, there are still a number of challenges for exploring large conformational changes in experiment. On the other hand, computational exploration of large conformational changes is made difficult by the fact that the time scale of the conformational changes are often on the order of sub-seconds to minutes, while the molecular dynamics simulations for moderately sized proteins are on the order of microseconds, thus

presenting a challenge for modeling conformational changes. Advancing the modeling in this would not only provide us the local information and connect the modeling directly to experiments, but would also predict and give guidance for the ongoing and future experimental explorations of global conformational changes.

In recent years, theoretical models have been developed to explore functional dynamics near the bottom of the energy landscape through integrating native structural information from two reference states. These models have been used to investigate the conformational transitions of typical allosteric proteins such as adenylate kinase (ADK) [4–8], Rop dimer [9], GFP [10], glutamine-binding protein [11], Arc repressor [12], calmodulin [13,14], Src kinase [15], NtrC [16] and protein kinase A [17] etc. With advanced experimental approaches, more and more systems have been revealed to have more than two states. While it is well established that the inactive ligand-free state and the active ligand-bound state are both critical to protein function, it is worth emphasizing that other metastable conformational states are also of importance. However, the present models with two basins seem to be insufficient to describe such multiple-state systems. Recently,

Author Summary

A central goal of biology is to understand the function of the organism and its constituent parts at each of its scales of complexity. Function at the molecular level is often realized by changes in conformation. Unfortunately, experimental explorations of global motions critical for functional conformational changes are still challenging. In the present work, we developed a coarse grained triple-well structure-based model to explore the underlying functional landscape of maltose-binding protein (MBP). By quantitative flux analysis, we uncover the underlying mechanism by which the major induced fit and minor population shift pathways co-exist. Though we have previously lent credence to the assertion that dynamical equilibrium between open and minor closed conformations exist for all the free PBPs, the generality of this rule is still a matter of open debate. We found that the hinge flexibility is favorable to population shift mechanism. This finding provides a theoretical explanation of the mechanism discrepancies in PBP protein family. We also simulated the folding dynamics using this functional multi-basin model which successfully reproduced earlier protein melting experiment. This represents an exciting opportunity to characterize the interplay between folding and function, which is a long-standing question in the community. The theoretical approach employed in this study is general and can be applied to other systems.

a multiple-basin landscape was explored from microscopic models [5] and macroscopic models [18,19]. In these models either the realization of macroscopic states is hard to control or the implementation is not readily available in common molecular modeling packages [19]. In this work, we developed a coarse grained triple-well structure-based model with the consideration of ligand binding, extending the two-state approach [7] for recapitulating multiple state large amplitude conformational changes.

We explored maltose binding protein (MBP) which has three observed states. MBP is a member of the large family of periplasmic binding proteins (PBPs) [20] with a similar (two-domain) folding structure and size, linked by a rather flexible β -strand region. Proteins in PBPs family display a common functional conformational switch between the apo-open state and the holo-closed state upon ligand binding. This is proposed to be a consequence of natural selection for fitting their function for efficient uptake in the periplasm and for directed chemotaxis [21]. A database of the PBP superfamily with hundreds of X-ray structures from *E. coli*, thermophilic bacteria and eukaryotes etc., is a treasure trove for studying the relationship between ligand binding and conformational coupling as well as for applications in protein engineering [20].

MBP is found in the periplasmic space of gram-negative bacteria and serves as a receptor for osmotic shock and chemotaxis in response to the maltose and other maltodextrins from the environment [22,23]. It has been extensively studied by X-ray crystallography [24–27], NMR spectroscopy [28–30], atomic force microscope [31] and other biophysical techniques [32,33]. Recently, NMR paramagnetic relaxation enhancement (PRE) studies [33] have revealed that a minor species of apo MBP which represents a partially closed state in solution without ligands. This is in addition to the major species and is consistent with the structure of open MBP resolved by X-ray. Presently, we face several mechanistic questions: What are the processes by which conformations change? Is the system better described by an induced-fit [3] or population shift mechanism [34]? How many

reaction pathways are there? What are their importance and the relative weights? Which part(s) of the protein plays the most important role? We addressed these questions using molecular dynamics simulations based on our structure based model.

So far, there are relatively few computational studies on the conformational dynamics of MBP [35–38]. Stockner et al. used short-time (30 ns) atomic MD simulation to investigate the functionally important transition of MBP [35]. They observed an open-to-closed transition starting from the open structure with ligand and a closed-to-open transition from close structure without ligand. However, the minor partially closed state was not found during the process, reflecting that the simulation time scale is insufficient. Most recently, the dynamical equilibrium between open and apo-closed states in free MBP was validated in MD simulations carried out by Bucher et al. [36,37] and Kondo et al. [38]. They explored the conformational dynamics by an enhanced sampling algorithm called accelerated MD [39], or by umbrella sampling to overcome the high free energy barriers. These methods both require one to modify the true potential by introducing a biasing potential into the system, making it difficult to obtain kinetic information. It is still challenging to draw definitive conclusions due to the computational bottleneck in the number of transition events.

In this work, we explored the multi-state dynamics from a global perspective by building a coarse-grained functional landscape based on available structural information (Figure 1), which was sampled sufficiently by molecular dynamics simulation. The present work allows us to perform a complete dynamical simulation of conformational transitions between all three observed states of MBP, and obtain thousands of transitions giving statistical reliability. In addition, this enables us to extract the thermodynamics and kinetic information directly from the equilibrium MD trajectories. Furthermore, these trajectories can be projected onto any arbitrary coordinates, such as in Cartesian, dihedral, or contact space as suggested in Ref. [40]. This feature is lack in the umbrella sampling approach or similar methods which require prior knowledge of reaction coordinates [39]. Overall, coarse grained models can provide a useful first order approximation and global landscape while the atomistic simulations explore the system in greater detail. The combination of both methods may provide a fuller picture of functional conformational transition. This combination strategy has been applied by us on binding-folding dynamics [41].

Based on our model, we approximated various perturbations including ligand concentration and hinge flexibility. We revealed the dependence of shape of the functional landscape and the robustness of the underlying mechanism of conformational transition in MBP. We established a link between the underlying energy landscape and thermodynamic stability, kinetic paths/speed, and structure function correlation.

Results

In this work we integrated structural information of three states of MBP to explore the conformational transition mechanisms. The key to construct a triple-basin model is to define a mixed contact map which integrates heterogeneous sources of structural information together. In our model, the mixed contact map contains the core contacts which are shared with all the native states as well as the state-specific contacts. The strengths of the core contacts and the state-specific contacts are rescaled by their energetic contributions which are denoted as ϵ_1 and ϵ_2 , respectively. In addition, the effect of ligand binding is realized by introduction of ligand mediated interactions, which are represented by inter-domain ligand-mediated

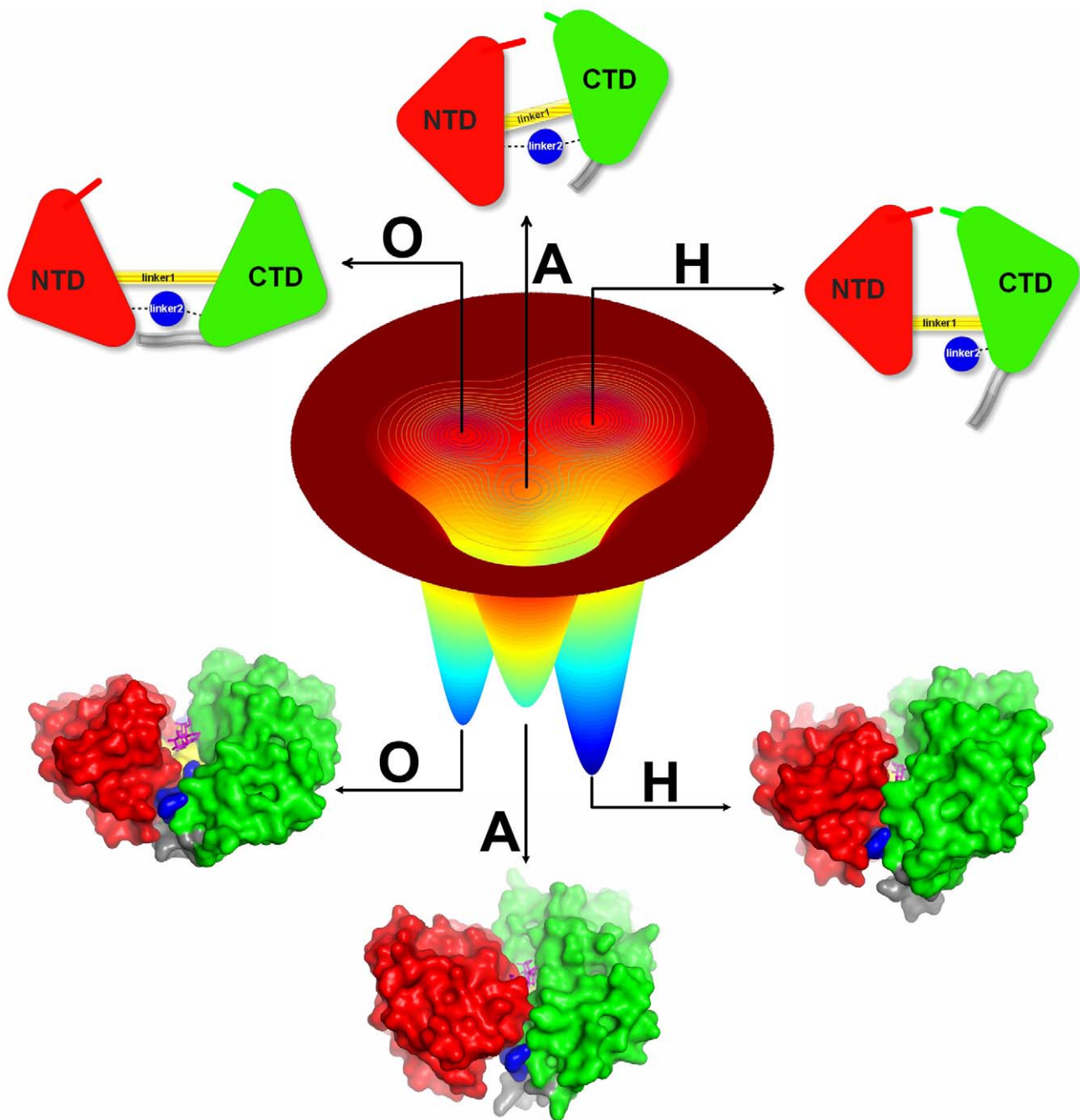


Figure 1. Schematic diagram of triple-basin landscape constructed by integrating structural information from open (O), apo-closed (A) and holo-closed (H) states of MBP. MBP contains two globular domains, NTD (colored in red) and CTD (colored in green) which can be further divided into C1 domain and C2 domain, sharing similar secondary structures and topology with a central β -sheet and on both sides with two or three parallel α -helices. These two large globular domains move back and forth by two linkers: linker1 (colored in yellow) and linker2 (colored in blue). The ligand-binding site is located at the base of a deep groove formed between the two domains. This functional interface, called the “balancing interface”, was considered to play a role as a “molecular switch” that triggers the conformational turnover. In fact, the balancing interface includes two important segments. One is a loop region (residues Y167 to D173, colored in grey) as a part of C2 domain, As MBP in open state, this loop interacts with NTD to form a number of contacts which are broken in holo-closed state or apo-closed state. Another is the linker2 (residues D310 to P330, colored in blue) which plays important roles not only in bridging the two domains, but also as the base of the active site groove. In this study, we refer to the loop as the “balancing loop” and the linker2 as the “balancing linker”.
doi:10.1371/journal.pcbi.1002471.g001

contacts formed in holo-closed state of MBP. The strength of ligand binding contacts is rescaled by ϵ_3 . See section of Model and Methods in *Text S1* for more details. In addition, molecular dynamics simulations were used to probe the sensitivity of the model on

energetic parameters. Several thousands of basin transitions indicated sufficient sampling.

Recently, a statistical survey for a set of allosteric proteins [42] revealed the robust allosteric feature that the state-specific contacts

are significantly weaker than the shared contacts. This feature is likely incorporated to facilitate conformational change by switching only the state-specific contacts. Considering this, the strength of state-specific contacts ϵ_2 is 0.4 and the strength of shared contacts ϵ_1 is 1.0. These parameters yield a ratio of state-specific contacts to shared contacts of 0.4. We found that the model worked well as the ratio was between 0.4 and 0.6. This indicates that the allosteric feature found in other allosteric proteins is also shared with MBP. To our knowledge, this work is the first explicit application of this allosteric feature to guide the parametrization of multiple-basin conformational change model.

Ligand-unbound functional landscapes

Based on dihedral angles parameterized from open state (Φ_O model, see section of Superior Angle Models in *Text S1* for details), we calibrated our triple-well model so that the relative population of open and apo-closed states is comparable to experimental measurements in the absence of ligand binding ($\epsilon_3 = 0.001$).

The free energy surfaces $F(Q_A, Q_O)$ is shown in the two-dimensional space of Q_A and Q_O for the ligand-unbound case (Figure S4 in *Text S1*). Here, Q_X and $RMSD_X$ (X can be O, A and H) are the order parameters which are employed to monitor the closeness to their respective states including open, apo-closed and holo-closed states (See the definitions in *Text S1*). $(Q_A, Q_O) = (0.9, 0.3)$ for the apo-closed conformation and $(0.1, 0.9)$ for the open conformation. The free energy barrier from apo-closed to open state is $6K_bT$, and from open to apo-closed state is $9K_bT$. Free energies are correlated with the equilibrium probabilities of the specific states which can be probed in biophysical tools such as spectroscopy, single molecule fluorescence [43], and atomic force microscopy (AFM) [44]. Equilibrium denaturation experiments [30] revealed a linear correlation between free energy of unfolding and the rotation angle between the two domains of MBP. The

stability of the (apo)protein decreases with domain closure by about $0.2Kcal/mol$ per degree of rotation. By measuring the difference in mechanical work between the ligand-free and ligand-bound states using AFM, the open-closed transition energy for MBPs was determined to be $4-12Kcal/mol$ [44]. From our simulation, we can conclude that, for free MBP, the free energy barrier from open state to holo-closed state should be larger than $10K_bT$, which is consistent with the experiment and provides the physical basis of the underlying energy landscape. We quantify the population of open and apo-closed state to be 94.0% and 4.9%, respectively, which agrees well with the paramagnetic relaxation enhancement data. Note that there is about 1.0% misfolded conformation located outside the native basins from the simulation.

Effects of ligand binding

Next, we investigate the effects of ligand binding by increasing ϵ_3 from 0.0 to 1.5. In Figure 2, the free energy surfaces $F(Q_A, Q_H)$ for different ligand-bound ratios are shown in the order parameter space of Q_A and Q_H . In Figure 2A, there are two free-energy minima: one corresponds to the open state at $(Q_A, Q_H) = (0.1, 0.6)$, the other corresponds to the apo-closed state at $(Q_A, Q_H) = (0.9, 0.6)$. It indicates that conformational transitions occur between the open state and the apo-closed state in the absence of ligand ($\epsilon_3 = 0.001$). In Figure 2B–E, a free-energy minimum corresponding to the holo-closed state at $(Q_A, Q_H) = (0.5, 0.9)$ connects the open basin and the apo-closed basin. This minimum does not exist in the absence of a ligand as shown in Figure 2A and the population of conformations at this minimum is small when binding to ligand with low concentration. In Figure 2F–H, we can see that the two basins for the open and the apo-closed states gradually disappear as the concentration of the ligands becomes high enough, making the holo-closed basin the unique one meaning that the protein stabilizes at the holo-closed state when saturated with bound ligand.

The effects of ligand binding

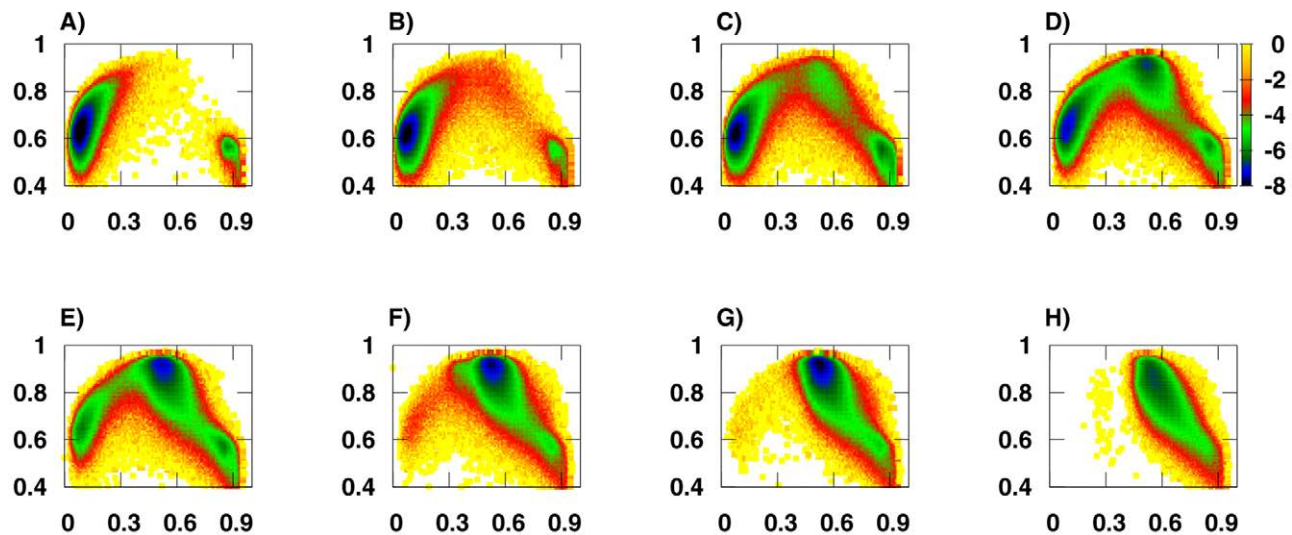


Figure 2. The effects of ligand binding on thermodynamics by increasing ϵ_3 from 0.0 to 1.5. The other parameters are $\epsilon_1 = 1.0, \epsilon_2 = 0.4, \epsilon_o = 1.0, \epsilon_a = 1.3, \epsilon_h = 0.2, \epsilon_p = 0.5, \epsilon_{hinge} = 1.0$. The free energy surfaces $F(Q_A, Q_H)$ are shown in the two-dimensional space of Q_A and Q_H for corresponding ϵ_3 values. (A) $\epsilon_3 = 0.001$, there are two free energy minima. One corresponds to open state at $(Q_A, Q_H) = (0.1, 0.6)$, the other corresponds to apo-closed state at $(Q_A, Q_H) = (0.9, 0.6)$. (B–E) $\epsilon_3 = 0.2, 0.3, 0.4, 0.5$, there is a free energy minimum corresponding to holo-closed state at $(Q_A, Q_H) = (0.5, 0.9)$ connects open basin and apo-closed basin. (F–H) $\epsilon_3 = 0.7, 0.8, 1.5$, the other two basins gradually disappear as the concentration of the ligands becomes high enough, making the holo-closed basin the unique one. doi:10.1371/journal.pcbi.1002471.g002

Induced fit pathways in free energy profiles

We further examine the underlying mechanism of how ligand binding and conformational change of MBP from the open state to the ligand-bound state are coupled. Two distinct mechanisms have been suggested to describe the conformational transition in biomolecular recognition. The first is called induced fit [3]. In this mechanism, the relationship between ligand binding and conformational change is accounted for as ligand binding drives a ligand-free, or apo (usually open) enzyme towards activated conformation. The second proposed mechanism is called conformational selection or population shift [34]. Within this paradigm, the unbound protein takes on multiple native conformations, subsequently, binding to ligand stabilizes the pre-existing higher energy conformation.

For the conformational change of MBP responding to ligand binding, the population shift mechanism indicates that ligand binding stabilizes the pre-existing partially closed state of MBP (A basin), and the protein changes its conformation into fully closed state (H basin) passing through A basin. The induced fit mechanism indicates the direct transitions from O basin to H basin without the need of passing through the minor populated state A. Therefore, the mechanism can be simply described by two possible kinetic routes: (i) the “induced fit route” (IF route), where predominantly the open form of MBP changes into H basin directly, that is, $O \rightarrow H$; and (ii) the “population shift route” (PS route), where MBP arrives at the H basin through the A basin, that is, $O \rightarrow A \rightarrow H$. Therefore, we can infer the mechanism of conformational change of MBP based on the state-transition routes under the free energy landscape. Rigorously speaking, the PS route represents that MBP first shifts the pre-existing distribution of conformations upon ligand binding followed by local conformational adjustment.

The conformational energy landscape of MBP shows that the landscape has three major basins with one basin biased towards the open and the other two toward the apo-closed and holo-closed conformations of the protein. This indicates that the experimentally determined structures are perfectly integrated into our model. The entire conformational dynamics of the system is characterized by the inter-basin transitions, which depends on the barriers between and relative energies of basins [5]. The conformational energy landscapes are reshaped by the change of ligand concentration which can modulate the height of free energy barriers or the depth of native state basins.

In Figure 3A–B, the typical time trajectories as a function of RMSDs ($RMSD_O, RMSD_H, RMSD_A$) and Q fractions (Q_O, Q_A) are shown. It indicates transitions between open and holo-closed states, and between holo-closed and apo-closed states. In addition, Figure 3C–D show the free energy profile as a function of Q_A and the two-dimensional free energy profile as a function of Q_A and Q_H . We found a direct route from open basin to holo-closed basin clearly implying an induced fit pathway. The apo-closed basin is not on the way to holo-closed basin, but is, in fact, circumvented en route.

Hidden population shift pathways revealed by kinetic analysis

By analysing the kinetic trajectories, we found the evidence that PS routes exist which are hidden in thermodynamic free energy profiles (Figure S5 in Text S1). A typical kinetic PS route in MBP is shown in Figure S6 in Text S1. A thermodynamically invisible path may actually exist. Under certain conditions related to reactant concentrations and rate constants, the flux through a given pathway can quantitatively reflect a reaction proceeding along the respective path [45].

From the energy landscape perspectives, the induced fit is a limiting case of conformational selection when the interaction partner selectively binds to the lowest energy conformation [3]. Hence, the actual mechanism in each allosteric protein may reside between the two limits. In many cases, the allosteric pathway may be modulated by the ligand and protein concentrations, so that both mechanisms may occur at certain conditions [46]. To precisely measure how much a transition proceeds through a given route, we calculate the reactive flux through that route. From our calculation, only $O \rightarrow H$ transition flux (F_{IF}) and $O \rightarrow A \rightarrow H$ transition flux (F_{PS}) are computed to analyse the relative importance of two limiting mechanisms when a ligand binds to MBP.

Furthermore, we investigated the dependence of fractional flux of IF route (f_I) and PS route (f_P) on the ligand concentration and hinge flexibility. Note that a comparison of the effects with different parameters should be performed under the same conditions. The fractional flux f_I and f_P can be calculated as follows:

$$f_I = \frac{F_{IF}}{F_{IF} + F_{PS}} = \frac{N_{O \rightarrow H}}{N_{O \rightarrow H} + N_{O \rightarrow A \rightarrow H}}$$

$$f_P = \frac{F_{PS}}{F_{IF} + F_{PS}} = 1 - f_I$$

where $N_{O \rightarrow H}$ is the number of direct transition from the open basin to the holo-closed basin without passing through any other basins. $N_{O \rightarrow A \rightarrow H}$ is the number of sequential transitions from the open basin to the apo-closed basin and then arriving at the holo-closed basin.

The basin dynamics of the system was investigated by applying the method analogous to the one described in [47] that explored the folding free energy landscape. We construct a reduced kinetic scheme by considering the transitions just between the native basins. The transition rate constant for this model, for example k_{OH} , is estimated as the ratio of transitions number $N_{O \rightarrow H}$ to total residence time in open basin. Generally, $k_{A \rightarrow B} = \frac{N_{A \rightarrow B}}{P_A * T_{tot}}$, where $N_{A \rightarrow B}$ is the the number of transitions from basin A to B, P_A is the probability of protein in basin A, T_{tot} is the duration of the trajectory. Note that $k_{A \rightarrow B}$ is in unit of ns^{-1} .

We next considered the kinetic response of MBP to changes in ligand concentration. The relationships of the fractional flux and transition rate constant with ϵ_3 are shown in Figure 4. From Figure 4A, it is clear that the IF transition flux increases with increasing ϵ_3 , especially for slightly higher values ($\epsilon_3 > 0.5$). The correlation between f_I and ϵ_3 indicates that the increasing ligand binding interactions encourage MBP to follow an induced fit pathway. This finding is consistent with the conclusions by Hammes et al. [45] that increasing ligand concentration favors the induced fit pathway. Intriguingly, f_I is always kept at relatively high values ($f_I > 0.7$) for all ligand concentrations. This strongly supports that the induced fit pathway is the predominant activation route of the system. The contribution of PS routes to the total activation flux does not exceed 35% (Figure S7 in Text S1). As pointed out by Zhou [46], the conformational transition rates are also a key factor in controlling the population of pathways. Figure 4B indicates that the transition rate constants k_{OH}, k_{OA} and k_{AH} increase as ligand concentration increases. However, k_{OH} increases more sharply than k_{OA} and k_{AH} . This can be easily understood considering that while the conformational transitions between the open and the holo-closed basin are

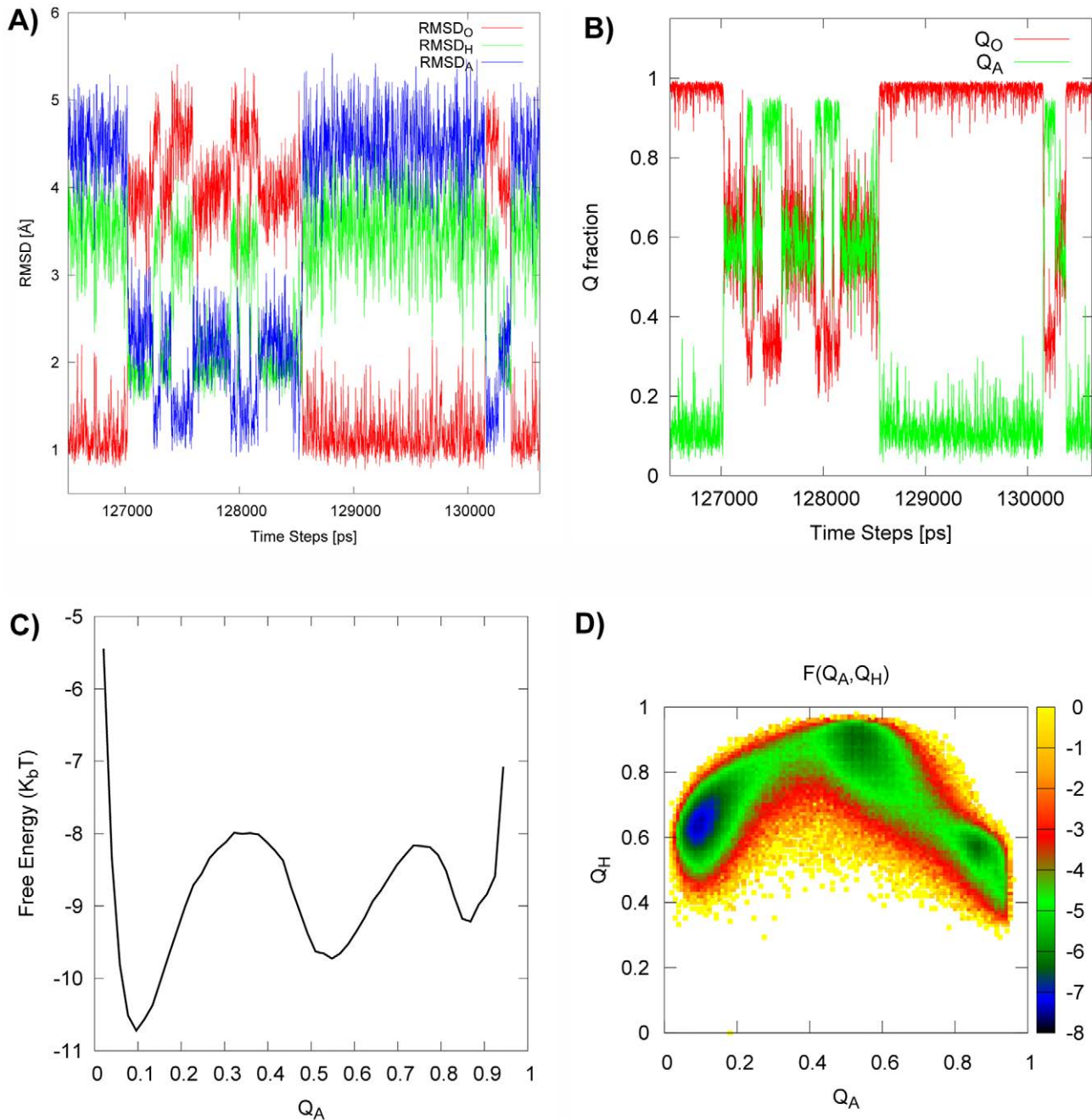


Figure 3. Induced fit pathway revealed by conformation switching between open, holo-closed and apo-closed states. The parameters are $\epsilon_1 = 1.0, \epsilon_2 = 0.4, \epsilon_o = 1.0, \epsilon_a = 1.3, \epsilon_h = 0.2, \epsilon_3 = 0.4, \epsilon_p = 0.5, \epsilon_{hinge} = 0.3$, using Φ_H potential. (A–B) shows the typical trajectories in time for Q fractions and RMSDs. (C) the free energy profile as a function of Q_A . (D) the two dimensional free energy profile as a function of Q_A and Q_H . There is a direct route from the open basin to the holo-closed basin implying a clear induced fit pathway. The apo-closed basin is not on the way to the holo-closed basin, but is, in fact, off pathway. In addition, we can identify two transition states TS_{OH}, TS_{AH} between O, A, H basins which are located at $(Q_A, Q_H) = (0.3, 0.8)$ and $(0.7, 0.65)$.
doi:10.1371/journal.pcbi.1002471.g003

fast, MBP can quickly switch to the active state directly from the open state. Our simulation results indicate that if the transition rate between the ligand-free major state and the ligand-binding active state is sufficiently high, then the protein will mostly follow the direct conformational transition route that results in a predominant induced fit mechanism.

The relationships between kinetic behaviors and ligand concentration were studied using Φ_H model (Figure 4). To alleviate the model dependence caused by the effects of the dihedral barrier of

the hinge regions in different dihedral models, we employed parameter $\epsilon_{hinge} = 0.01$. Note that ϵ_{hinge} is introduced to rescale the energetic contribution of the angle term for the hinge regions. Thus, this parameter can be used to tune the hinge flexibility. The effects of the hinge flexibility on kinetics are discussed in detail in the following section.

In particular, we analyze the detailed kinetic trajectories generated by the specific model with Φ_H potential using a parameter set with moderate values ($\epsilon_3 = 0.4, \epsilon_p = 0.5, \epsilon_{hinge} = 0.3$).

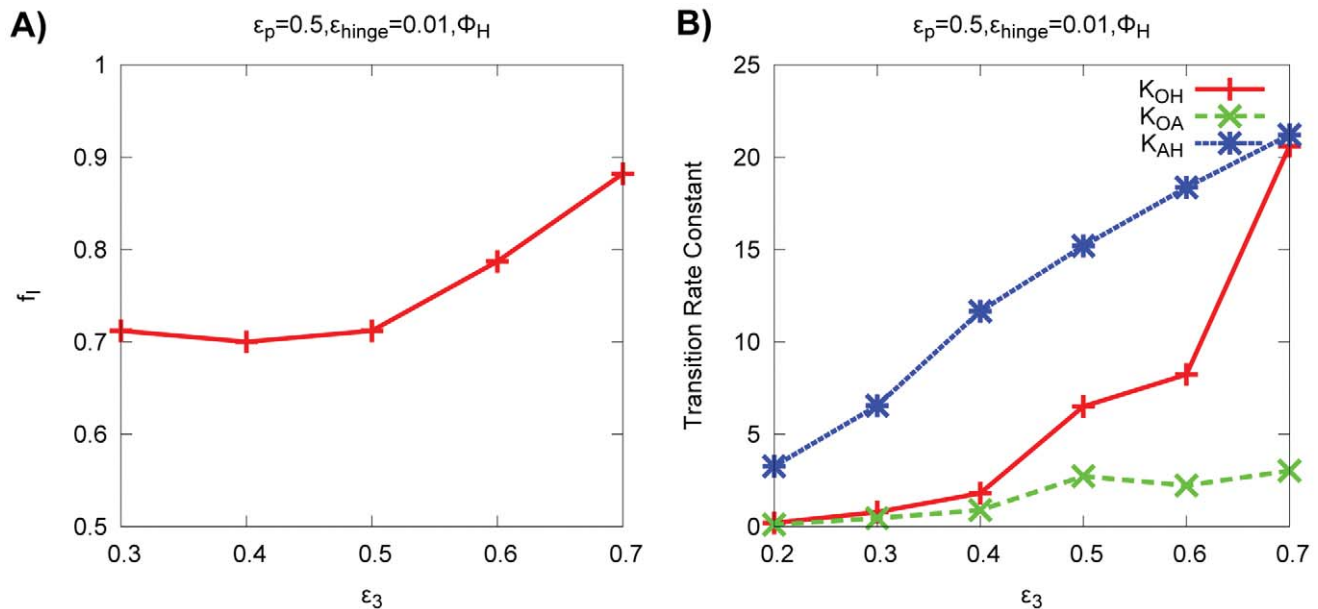


Figure 4. Influence of ligand concentration on kinetic mechanism and transition rate constant. (A) Fractional flux of IF route (f_I) as a function of ligand concentration. Increasing ligand binding interactions can facilitate MBP to activate its conformation following an induced fit pathway. Intriguingly, f_I is always kept on high values ($f_I > 0.7$) during all the ligand concentrations. This strongly supports that the induced fit pathway is the predominant activation route of the system. (B) Transition rate constants k_{OH} , k_{OA} , k_{AH} as a function of ligand concentration. k_{OH} , k_{OA} and k_{AH} increased as ligand concentration increased. However, k_{OH} increased more sharply than k_{OA} and k_{AH} . Our simulation results give strong support that if the transition rate between ligand-free major state and ligand-binding active state is sufficiently high, then the enzyme will mostly follow the direct conformational transition route, resulting in a predominant induced fit mechanism.
doi:10.1371/journal.pcbi.1002471.g004

Note that the corresponding free energy profiles under the same model are shown in Figure 3. Figure 5A shows the schematic representation of the basins dynamics. The transitions between the basins are represented by arrows along with the transition numbers. Transitions that occur clock-wise are represented as blue arrows. The fractional flux of IF route is 0.796. It means that 79.6% of open to holo-closed transitions proceed through induced fit pathway, the rest follow population shift pathway. The basin probabilities P_O , P_A and P_H are 51.2%, 33.1% and 15.0% respectively. The percentages of basin probability do not necessarily add up to exactly 100% because of minor misfolding conformations outside these native basins.

Moreover, for the timescale of the transitions between the open state and the apo-closed state, the experiments suggest an upper limit of 20 μ s [48] and a lower limit of 20 ns [33]. In addition, the O \rightarrow H transition rate can be estimated around 30–50 ns [36] by experimental association rate constants that fall within a very narrow range of $2-3 \times 10^7 M^{-1}s^{-1}$ [49], despite the differences in the size and nature (linear and cyclic) of the maltodextrin ligands. In our simulations, the activation transitions from ligand-free MBP is illustrated in Figure 5B in which k_{OH} , k_{OA} and k_{AH} are 2.9, 0.9, 13.0 ns^{-1} , respectively. It should be noted that, in our coarse grained structure-based model, transitions between native basins are significantly faster than the realistic time scale. The absolute time scales cannot be obtained due to the coarse grained nature of our model and the lack of explicit solvent molecules. It is, however, possible to make an order of magnitude estimate based on the gap of time scale between simulations and experiments. Accordingly, it may be reasonable to scale the simulation time by two or three order of magnitude. This transformation should be taken as a very crude order of magnitude estimate.

Clearly, the rate of A \rightarrow H transition is about 4 times the O \rightarrow H transition rate, and 15 times the O \rightarrow A transition rate. The finding

that k_{OH} is larger than k_{OA} in our model is supported by the fact that the open to holo-closed transition can be observed in MD simulations that are at least 30 ns long [35], and the pre-existing equilibrium in apo MBP between an open and a partially closed conformer was only observed after 300 ns of simulation time [36]. The true transition rates should be dependent on a number of factors, such as temperature, pH, and ligand concentration. Here, they are calculated under a condition corresponding to modest ligand concentration ($\epsilon_3 = 0.4$). The results indicate that the relatively less contribution of PS routes is due to the fact that O \rightarrow A transition is the rate-limiting step, although the protein transitions into the H basin from the A basin more quickly than from the O basin.

Influence of hinge flexibility on thermodynamics and kinetics

It is widely accepted that enzyme activity is closely related to the fast, local fluctuations reflecting the flexibility of mostly the hinge regions. These hinges were computationally identified from differences in both pseudo-angles and pseudo-dihedral angles between native states (see details in section Model and Methods in Text S1). In the following, we will investigate the influence of hinge flexibility on the thermodynamics and kinetics of the system.

First, the influences of hinge flexibility on thermodynamics are summarized in Figure S8 in Text S1. Considering the different native angle biases of Φ_O and Φ_H model, we compared the results to assess the model dependence. With Φ_O model, the hinge flexibility (the decrease of ϵ_{hinge}) is found to mainly decrease the depth of the open basin, by destabilizing the open-closed basin and increasing the stability of the apo and holo-closed basins, but has little influence on either H \rightarrow A or A \rightarrow H transition barrier. With Φ_H model, it shows that hinge flexibility mostly decreases the stability of holo-close basin and increases the stability of apo-closed

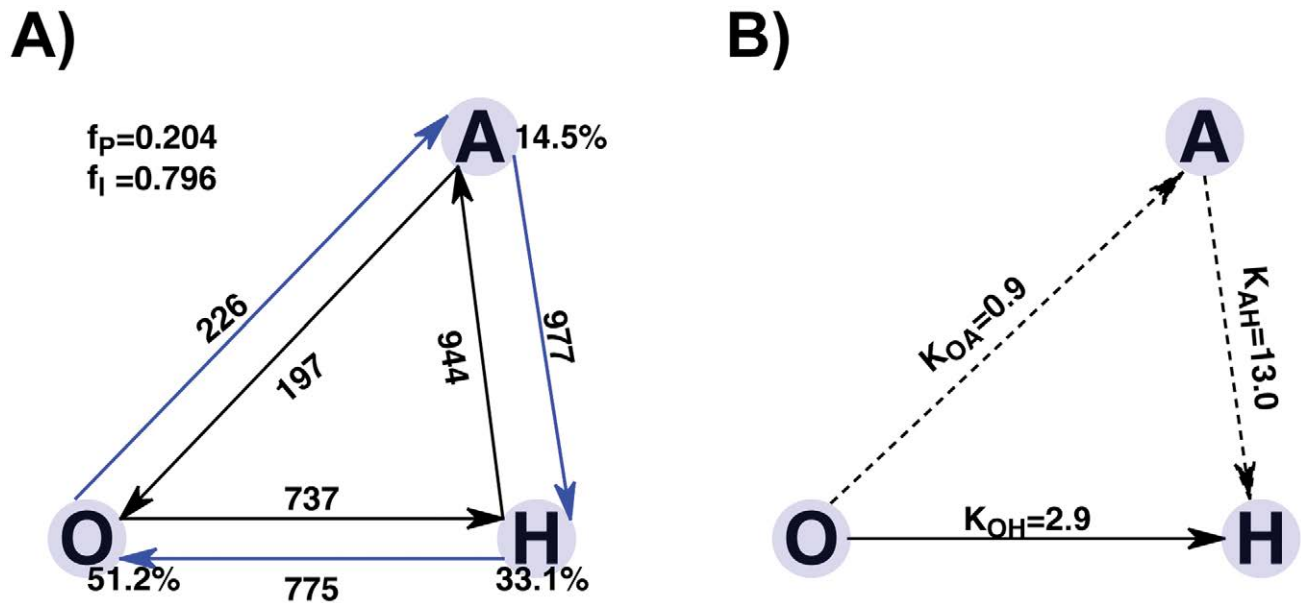


Figure 5. Schematic representation of the basins dynamics. $\epsilon_3 = 0.4$ and Φ_H model is used. The other parameters are $\epsilon_1 = 1.0, \epsilon_2 = 0.4, \epsilon_o = 1.0, \epsilon_a = 1.3, \epsilon_h = 0.2, \epsilon_p = 0.5, \epsilon_{hinge} = 0.3$. Open, apo-closed and holo-closed states are labeled by O, A and H respectively. (A) Transitions between these basins are represented by arrows along with transition numbers. Transitions that occur clockwise are represented as blue arrows. The basin probability P_O, P_A and P_H are 51.2%, 33.1% and 14.5% respectively. The fractional flux of the IF route is 0.796. It means that there are 79.6% of open to holo-closed transitions that proceed through the induced fit pathway, whereas the rest of the 20.4% of transitions follow population shift pathway. (B) Rate constants (k_{OH}, k_{OA}, k_{AH} in unit of ns^{-1}) of conformational change are shown. Direct O-H transition is represented by solid arrow, O-A and A-H transitions by dashed arrows. k_{OH}, k_{OA} and k_{AH} are 2.9, 0.9, 13.0, respectively. The rate of A-H transition is about 4 times the O-H transition rate, and 15 times the O-A transition rate.
doi:10.1371/journal.pcbi.1002471.g005

basin. However, there is little impact on $H \rightarrow O$ free energy barrier and stability of open basin. Although the models may have different native biases, they all imply that increasing hinge flexibility can decrease the free energy barriers from their native basin to other basins.

Furthermore, we investigate the effects of hinge flexibility on kinetics (Figure S7 in Text S1). From the correlations between fractional flux of IF routes and hinge flexibility, we can see that the increase of hinge rigidity in both models facilitates the conformational transition of MBP along the induced fit pathway. The PS routes are more favored for high hinge flexibility as the transition between the open and the apo-closed conformations is not only accompanied by domain rotation, but also by significant domain twist, both of which depend on hinge flexibility.

From the relationships between transition rate constants and hinge flexibility, we can see that in Φ_O model, increase of hinge flexibility can greatly accelerate the $O \rightarrow H$ transition. However the corresponding fractional flux f_I decreases as hinge flexibility increases. This can be explained as the competition between IF transitions and PS transitions as k_{OA} and k_{AH} also increase. To be more exact, k_{OA} increases 17.3 times and k_{OH} increases 10.6 times from $\epsilon_{hinge} = 1.0$ to $\epsilon_{hinge} = 0.0$. In contrast, in Φ_H model, the $O \rightarrow H$ transition process is slightly slowed down when increasing the hinge flexibility, and the corresponding fractional flux also decreases.

Overall, the flexibility of hinge regions play an important role in both basin stability and basin dynamics.

Structural characterization of transition states

The above thermodynamics and kinetics results reveal a mixed mechanism for the conformational change in MBP upon ligand binding. It consists of a major induced fit route and a minor

population shift route. We located the transition state by finding the extremum between the minima of the free energy profile as the saddle point of the free energy landscape. From the thermodynamic free energy profiles in Figure 3D, we can identify two transition states TS_{OH}, TS_{AH} between the O, A, H basins which are located at $(Q_A, Q_H) = (0.3, 0.8)$ and $(0.7, 0.65)$. And from Figure S4 in Text S1, another transition state TS_{OA} between the O and A basin can be identified and located at $(Q_A, Q_O) = (0.6, 0.5)$. First, we describe the structural information in the transition state ensembles.

In Figure 6 and 7, we show the contact probability maps for transition state ensembles of TS_{OH} and TS_{OA} . We map these contacts onto the three-dimensional structures of MBP. Note that only the state-specific contacts are illustrated. This is due to the fact that most of the contacts are shared by the two ending transition basins. Their probabilities are shown by points with different colors according to side color bar in which dark colors mean high probability and light colors mean low probability.

Figure 6 shows the structural characterization of the transition state between the open and holo-closed states. Clearly, we can see that the contacts between the tip of balancing loop and NTD (C region) are broken and the interface between NTD and CTD at the tip of ligand binding cleft (D region) forms 7 native H-specific contacts (70% formation). Note that H-specific contacts represent the contacts that are specific for the holo-closed state. A-specific denotes contacts for the apo-closed state and O-specific denotes contacts for the open state. By contrast, the interface at the tail of cleft (E region) only forms 2 contacts (25% formation), and the long-range H-specific contacts whose optimal distances in open state are larger than 20 Å are not formed. However, the contacts between balancing the linker and NTD (A region) and part of contacts between balancing the loop and NTD (B region) are not

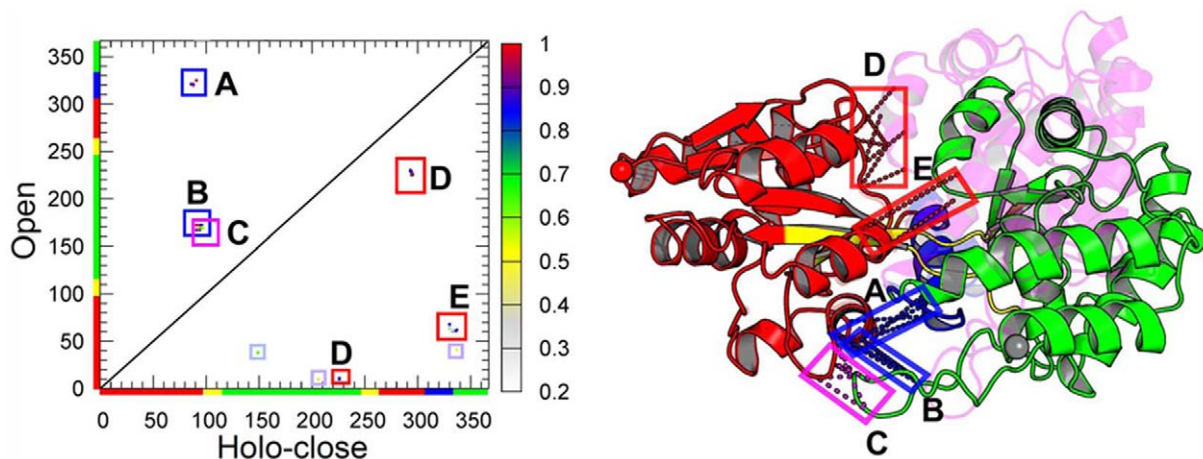


Figure 6. Structural characterization of transition state ensembles TS_{OH} . Native state-specific contacts with different probability formed in TS_{OH} are mapped onto the open and holo-closed crystal structures. In the left column, it shows the residues' contact probability maps. Note that, we only show the state-specific contacts because most of contacts are shared by the two-ending transition basins. The upper triangle (lower triangle) corresponds to probability map of the O-specific (H-specific) native contacts. The corresponding probability for a particular residues' pair forming two-body native contact is illustrated according to the side color bar in which dark colors mean high formation probability and light colors mean low formation probability. The segments corresponding to the NTD (1–94 and 259–310, red), CTD (113–250 and 331–366, green), linker1 (102–112 and 250–260, yellow) and linker2 (310–330, blue) are labelled in the X-axis (holo-closed state) and Y-axis (open state) with different colors corresponding to the structures of the right column. In the right column, the state-specific contacts are mapped onto the three-dimensional structures of MBP. Note that the NTD is used to superimpose open and closed conformations, and the CTD for holo-closed structure is represented by transparent cartoon in red. Red squares (D and E regions) and magenta squares (C regions) highlight the H-specific interactions formed and the O-specific interactions broken in TS_{OH} respectively. Blue squares (A and B regions) label the O-specific interactions that are not broken. The H-specific contacts not formed in TS are labeled by grey squares. They are not mapped onto the structures for clarity.
doi:10.1371/journal.pcbi.1002471.g006

broken. We propose that the balancing loop movement switches the closure of the tip of ligand binding cleft and the balancing linker motion triggers the final closure of the cleft in open to holo-closed transition.

Figure 7, shows the structural characterization of TS_{OA} . We can see that contacts (C region) between the balancing loop and NTD are broken and a number of contacts (D region) between the balancing linker (mostly located at residues K322 to I329) and NTD are also broken. The interface at the tip of cleft (E region) forms all A-specific native contacts (100% formation). A part of A-specific contacts at tail interface (F region) is formed and the rest (G region) is not formed. In addition, the O-specific contacts between helix part of balancing linker and two domains (region A and B) are not broken. Comparing TS_{OA} with TS_{OH} , we found that more contacts between the balancing interface (containing the balancing loop and balancing linker) and NTD are broken in TS_{OA} , and more contacts are formed between the interface at the ligand binding cleft. It seems that not only the balancing loop acts as a switch that initiates the domain closure, but also the balancing linker plays an important role in the conformational change whose role was not investigated in the earlier simulation [36].

Note that A-specific contacts in H-region contain pairs D10-E107, K11-E107, V106-Q321, V257-Q321 which represent the interactions between linker1 and NTD and between linker1 and linker2 in apo-closed state. These contacts are not formed in TS. The green squares in contact probability map and green dashed lines in mapped structure are used to label the O-specific contacts (F88-V255 and V106-A299) between linker1 and NTD at the bottom of the ligand bind cleft, which are not broken in TS. It indicates that the domain twist does not occur in TS_{OA} because that the twist has to disrupt the O-specific contacts F88-V255 and V106-A299 and may form A-specific contacts (in H region) between linker1 and the loop1 (N8-K11) in NTD (such as, ion pair K11-E107). From this point, we propose that domain partial

closure occurs before domain twist in the transition between the open and apo-closed states.

Functional Φ values and local cracking in transition states analysis

Characterization of the transition state properties, especially the inhomogeneous distribution of contacts between residues, will help us to understand the microscopic structural mechanism of conformational change through locating the sites or seeds for the nucleation. The ϕ values provide important characterization for particular residues at the transition state ensemble. By experimental Φ value analysis, one can identify the critical residues with high values that cluster together in the transition state to form the nucleus for conformational change from the unfolded state to the native folded state. In our model, we use functional Φ values to characterize the transition state ensemble for conformational changes between multiple basins [6]. See details of functional Φ values calculation in *Text S1*.

We calculate the functional Φ_F -values $\Phi_{F_i}^{OH}$ and $\Phi_{F_i}^{OA}$ for particular residues with a sensible difference in the thermal mean number of interaction contacts in the transition states of $O \leftrightarrow H$ transition and $O \leftrightarrow A$ transition. The results are summarized in Figure S9 in *Text S1*.

It shows that certain residues in NTD (D10, K11, K293, P294, L295) and in CTD (P225, W226, S229) play a key role in stabilizing the holo-closed state ($\langle Q_i \rangle^H$ is larger than $\langle Q_i \rangle^O$) and contribute to the closure of MBP ($\Phi_{F_i, OH} > 0.6$). These residues are clustered together in the holo-closed state to form a tightly interacting network corresponding to the high contact probability region in Figure 7. Our model also predicts that the $\langle Q_i \rangle$ values of residues F88, D91, R94, N96, Y167, N169, Y172 and Q321 in TS_{OH} are especially close to their $\langle Q_i \rangle$ in open state ($\Phi_{F_i, OH} < 0.2$). This can also be explained by the contact probability map where O-specific contacts (A region) between

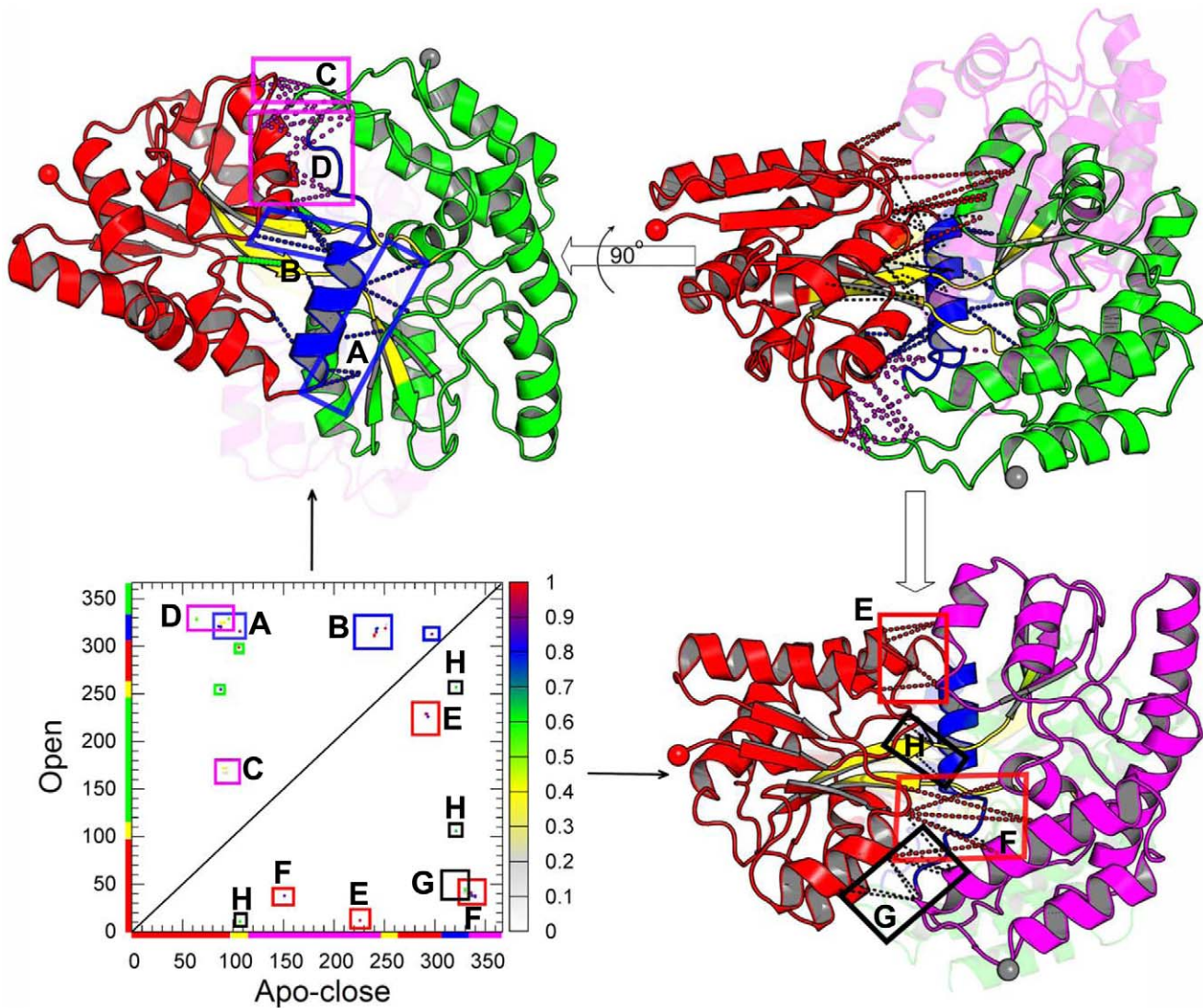


Figure 7. Structural characterization of transition state ensembles TS_{OA} . Native state-specific contacts with different probability formed in TS_{OA} are mapped onto the open and apo-closed structures (upper right subfigure). For the contact probability map (lower left subfigure), the upper triangle and lower triangle correspond to probability map of the O-specific and A-specific native contacts, respectively. Note that the NTD is used to superimpose the open and closed conformations. For clarity, the O-specific mapped structures and A-specific mapped structures are shown separately in upper left and lower right corners respectively. For O-specific mapped structures (upper left subfigure), which are turned 90° for the illustration of the back of active site groove, the CTD for apo-closed structure is represented by transparent cartoon in red. For the A-specific mapped structures (lower right subfigure), the CTD for the open structure is represented by the transparent cartoon in green. Blue squares (A and B regions) label the O-specific interactions which are not broken in TS_{OA} . These broken O-specific interactions are highlighted by magenta squares (C and D regions). Red squares (E and F regions) and black squares (G and H regions) highlight the formed and unformed A-specific interactions respectively. The green squares in the contact probability map and green dashed lines in the mapped structure are used to label the O-specific contacts (F88-V255 and V106-A299) between linker1 and NTD at the bottom of the ligand bind cleft, which are not broken in TS.

doi:10.1371/journal.pcbi.1002471.g007

balancing linker and NTD and partial O-specific contacts (B region) between balancing loop and NTD are intact. In addition, the Φ_{F_iOH} values for residues Y95, N96, G170, K171 (white spheres) are about 0.5 and $\langle Q_i \rangle^O$ values for them are larger than $\langle Q_i \rangle^H$, indicating that the interactions between the balancing loop and NTD are partially broken in TS_{OH} . This is consistent with the result reported from contact probability maps.

The TS_{OA} analysis indicates that the clustered residues are the same as in TS_{OH} except that K38 also plays an important role in stabilizing the apo-closed state. In addition, the Φ_{F_iOH} value for E107 is 0.7, which indicates that E107 also plays a favorable role in stabilizing the apo-closed conformation in TS_{OA} . These

interactions may include the salt bridge between K11 and E107 as a member of the “hook-and-eye” motif, which can be important to lock the protein-ligand complex in a semiclosed conformation [35]. E107 is located at the base of the binding cleft, in the middle of linker1. Compared to the open state, the side-chain carboxylate C_δ shifts 3.4 Å in ligand-bound closed state and 6.2 Å in apo-closed state thus moving up into the cleft. Furthermore, for helix A15 in C2 domain (see secondary structural definition in *Text S1*), there are not only high Φ_{F_i} residues (W336, Y337, R340), but also low a Φ_{F_i} residue (M332). This is consistent with the contact probability map in Figure 7 that shows part of native contacts in the apo-closed state is formed and part of them is absent in the

interface at the tail of the ligand binding cleft. The high Φ_{Fi} values for residues in the balancing loop (N169 and G170) indicates that the interactions between NTD and the balancing loop are completely broken as $\langle Q_i \rangle^A$ are close to 0 for F88, D91, N96, G97, N169 and G170.

Interestingly, we found that Φ_{Fi}^{OA} values of K252, K322 and G323 are larger than 1 (yellow spheres) as $\langle Q_i \rangle^{OA}$ is out of the range of $\langle Q_i \rangle^O$ and $\langle Q_i \rangle^A$. For K252, $\langle Q_i \rangle^O$, $\langle Q_i \rangle^{OA}$ and $\langle Q_i \rangle^A$ are 3.8, 2.8, 3.3 respectively. For K322, they are 3.9, 2.6, 3.1, and 2.7, 1.6, 2.2 for G323. K252 is located at the loop region of linker1, and K322 and G323 are located at the loop region of linker2 (balancing linker). Their abnormal Φ_{Fi} values indicate that the movements of linker1 and linker2 are not completely coupled during the open to apo-closed transition. Especially, for the loop region in linker2, the large structural fluctuation may cause it to move away from linker1, resulting in loss of contacts between K252 and K322 and G323. On the other hand, the loss of contacts in linkers brings about more flexibility for rigid-body domain movement. Such behaviour of breaking and reforming native contacts is known as “backtracking”. It is somewhat surprising that such a “backtracking” phenomenon is observed during the exchanging process of predominantly open form and minor partially closed form in MBP. To our knowledge, it is the first reported case in protein allostery, although such behavior has been observed in protein folding [50–52] and binding [41]. Such functional “backtracking” behavior is in favor of conformational change.

It is well known that large structural arrangement of protein can be realized by changes of several dihedral angles at the hinge regions. In some cases the high strain energy may accumulate in various localized regions. It can be relaxed by local unfolding or cracking if the energy exceeds a threshold [7]. See the detailed calculation of local unfolding in *Text S1*. The results of local cracking calculation are summarized in Figure S10 in *Text S1*. Our model indicates that the local unfolding points contain V106, L117, P119, E149, E168 and N328 in TS_{OH} . And for TS_{OA} , these unfolding regions include H60, I104, P150, Q148, T204, P244, K322, G323, M326. These residues are located at the loop regions with high flexibility with the exception of I104 and V106 approaching to the known hinge residue E107. They are located at linker1. The high C_i value (see definition in section of Local Cracking in *Text S1*) for E168 which is located at the balancing loop may be caused by the high flexibility of balancing loop that are not in contact with NTD. In addition, N328 in TS_{OH} and K322, G323, M326 in TS_{OA} are located at the loop region of the balancing linker.

In MBP, the hydrophobic residues have been found to be important to the domain closure [30,32,36], which makes the protein more compact by reducing the solvent accessibility. When the balancing interface loses contacts with NTD, the side-chain of non-polar residues become more solvent-exposed, which in turn drives a distant conformational change in active site cleft. More specifically, our simulation reveals that the solvent exposure of the balancing loop drives the tip of active site cleft to close. We also found that the charged residues are important, however, compared to TS_{OH} , the electrostatic interactions seem to be more important for TS_{OA} in which four charged residues (D37, K38, E41, R340) take part in the contact networks formed at the tail of active site cleft. Among these contacts, there is a salt bridge D37-R340 formed between NTD and CTD. Another inter-domain salt bridge K38-E149 can be found specifically in the holo-closed state, but it is not formed at TS_{OH} . In addition, the functional Φ_i value analysis reveals that E107 may help stabilizing the TS_{OA} through the formation of salt bridge between K11 and E107. Taken

together, we propose that the population shift pathway is more electrostatically driven than the induced fit pathway for MBP.

Dependence of MBP unfolding on ligand binding

We further explored the relationship between unfolding characteristics of MBP and ligand concentration through simulation. The thermodynamics of unfolding in MBP has been previously explored by a combination of differential scanning and titration calorimetry and fluorescence spectroscopy under different solvent conditions [53]. In this work, we simulated the folding dynamics using our functional model. The results are summarized in Figure S11 in *Text S1*. It indicates that the specific heat curves shift to higher temperatures with increasing ligand concentration, suggesting an increase of protein stability. In addition, the folding cooperativity increases as well, which is illustrated by the narrowing of the specific heat profile.

Structurally, MBP contains two globular domains. Between the two domains, it forms a deep groove whose base locates the ligand-binding site. The X-ray structure of MBP in complex with maltotriose [24,26,54] shows that the binding pocket is lined with a number of polar and aromatic groups from both domains that participate in hydrogen-bonding and van der Waals interactions with the ligand. These interactions make the two interlaced domains more tightly packed against each other. Overall, our simulations support that the presence of a ligand increases the cooperativity and stability of MBP because the ligand binding bridges the two domains more tightly. Higher ligand concentration or binding with ligands with higher affinity, makes the protein not only more stable, but more cooperative. This conclusion is in agreement with earlier experimental findings [53].

Discussion

Conformational transitions are central to a multitude of physiological processes, such as enzyme catalysis, and also essential for nonenzymatic binding events. Due to the limit of current experimental techniques, the molecular details for the microscopic structural mechanism of conformational change are often lacking. Therefore we developed a theoretical model for describing the structure-function relationship, and extended the double-well model [5–10,12–17], into a triple-well model, which we used to account for conformational switching between the open, apo-closed and holo-closed states of MBP.

It is worthwhile noting that the microscopic mixed double-well models with two reference conformations may construct additional basins in the free energy landscape, such as in Lu and Wang’s microscopic double-well model [5] and in Whitford’s mixed contact map model [6]. However, these models don’t necessarily yield the required free energy basins. For example, Lu and Wang’s microscopic double-well model can generate four free energy basins for ADK [5], but no additional basins beside the two reference basins for glutamine-binding protein (GlnBP) [55]. In fact, we have attempted to build a dual-basin model, which only integrated the open and the holo-closed conformations of MBP. However, this model failed to generate the apo-closed basin, demonstrating the limitations of these double-well models in simulating complex multi-state systems. Such limitations are what led us to develop the present triple-basin model.

Constant temperature molecular dynamics simulations were performed to obtain the free energy landscape and kinetics for conformational changes. It clearly reveals an induced fit pathway on the free energy surface. In addition, a minor population shift pathway is found through the analysis of kinetic trajectories. By introducing the flux analysis as suggested by Hammes et al. [45],

we measured the relative weight of the two parallel activation paths in a quantitative way. The results support a mixed mechanism of MBP in the presence of moderate concentration of ligand whereby major induced fit pathways and minor population shift pathways coexist.

The thermodynamic and kinetic manifestations of function transitions are closely related due to the correlation between transition rates and free energy barriers of basin hopping. Rigorously speaking, the population shift route ($O \rightarrow A \rightarrow H$) represents the population shift followed by local conformational adjustment (induced fit) similar to some other systems [16,56,57]. The mixed mechanism is consistent with the free energy landscape sampled by atomic simulation [37,38], where the ligand binding induced conformational change is proposed to be dominant for MBP. In our work, we quantified the relative weights of both induced fit and population shift from kinetic pathway perspectives. These results support the view that the underlying mechanism of conformational transition does not necessarily follow the population shift scenario, and that even free protein populates a wide range of conformations.

In nature, conformational change of protein upon ligand binding can be explained by either one of the two mechanisms or their combinations. It is dependent on the systems under different conditions. The relative weights of both pathways are dependent on a number of factors, including the ligand and protein concentrations, the conformational transition rates, and the properties of ligands etc. [11,45,46,58]. Other methods also have been developed, such as Markov state models (MSM) [57]. Our work provides a good example of quantifying the relative contributions of multiple possible mechanisms.

A “Venus Fly-trap” model was employed to explain the conformational change of PBPs family [59]. The hinge regions are believed to play a key role in control of the rapid response of PBPs upon ligand binding resembling that the carnivorous plant traps its prey [60]. In the present work, we investigate the role of the hinge in conformational change, especially for MBP. We found the flexibility of hinge regions plays an important role not only in the basin stability and but also in basin dynamics. The results also support that increasing the flexibility of the hinge regions promotes the population shift route.

It is still a matter of debate whether it is a general rule that dynamical equilibrium between the open and the minor closed conformation exists for all the free PBPs though our current and past work certainly supports this assertion [33,37,38,57,61], although Bermejo et al. claimed that there was no such apo-closed state for GlnBP [61]. The findings in our simulation support the explanation that the population shift mechanism is unfavourable for conformational change of GlnBP due to the fact that its hinge relatively rigidified by strong hydrogen bond interactions.

Protein folding is critical for three dimensional structural formation, and the conformational switching between multiple well defined states is vital for interactions and functions at the molecular level. They are often studied separately. So far, there have been abundant studies to explore the structure-folding relationship [9,62]. Recently, more and more attention has been turning to the functional transitions and native dynamics of proteins [5,6,17,34,38]. However, there are very few studies to investigate the interplay between protein structure, folding, and function as a whole. In this work, we explored the folding landscape and functional landscape of MBP under the same theoretical framework using our triple-well structure-based model. To our knowledge, this work is the first study to provide such a description. In addition, the effects of ligand concentration in protein melting were also investigated. The conclusion that the

presence of ligand increasing the cooperativity and stability of MBP agrees well with earlier experimental findings. The success of reproducing experimental folding properties also implies that the coarse-grained implicit ligand model has the ability to capture the essential effects of ligand binding not only on functionally conformational change but also on the global folding landscape.

From the perspective of energy landscape theory, protein folding and protein function share a common funneled energy landscape [63] (Figure 8A). Proteins fold their conformation into low energy ensembles from the top of the funnel to the bottom, and carry out the biological function by changing their conformations between a modest number of folded states in response to ligand binding or environmental changes. Compared with the whole folding phase space, the number of conformations is much smaller in the functional landscape. It makes the bottom of the funnel more easily sampled using current computational resources, especially with the aid of coarse grained models. We can consider protein folding as the large-scale (inter-basin between native and non-native) conformational changes between unfolding state and folded states, and functional transitions as small-scale (intra-basin, native) conformational changes between multiple folded basins. Although the folding dynamics were usually sampled by structure-based models with a single native basin, in principle, these multi-basin models were specially developed for functional dynamics should have the ability to capture the folding dynamics, as done in the present work.

Furthermore, the physical reason why unfolding/melting occurs in the functional study can be answered using the the thermal energy scale (Figure 8B). When the temperature is high (in fact melting temperature is much higher than the functional transition temperature), the energy scale $K_b T$ is high. At this scale, the differences between global energy landscape of double well and triple well for function and single well funnel for folding are blurry and even indistinguishable. Conformational dynamics at this temperature in these models all look like large-scale folding/unfolding transitions between the top of the energy funnel and the bottom (native basins). At the low temperatures, corresponding to the low energy scale near native states, the conformational dynamics occur at the bottom of landscape (functional landscape), the difference between folding funnel model and functional multi-basin model will be explicit and distinct. In other words, on the coarsened level, we have a funneled energy landscape for folding. If we dive into the smaller scales of the energy landscape near the native states, we see the more detailed structure and functional energy landscape emerging (Figure 8C).

There have been several theoretical studies about conformational dynamics of MBP, most recently [36–38,64] stimulated by the experimental finding of hidden partially closed conformation [33]. There are substantial differences between these studies and the present work. Our model is based on energy landscape theory in which proteins minimize the conflicting interactions so as to be minimally frustrated and have globally funneled energy landscapes biasing towards their native structures as a consequence of evolution [34]. It is well established that the folding mechanism and binding mechanism are encoded in the topology of proteins in the native folded state [9]. Therefore, our model is developed based on the assumption that the functional transitions are mostly governed by protein topology. Our model is structure based, which can give both statistically reliable thermodynamic and kinetic information at the coarse grained level, while others that are more detailed atomic based giving statistically sampling limited thermodynamic information but no kinetic information.

Although, electrostatic interactions are taken into account in our model, it may be true that other energetic frustrations also

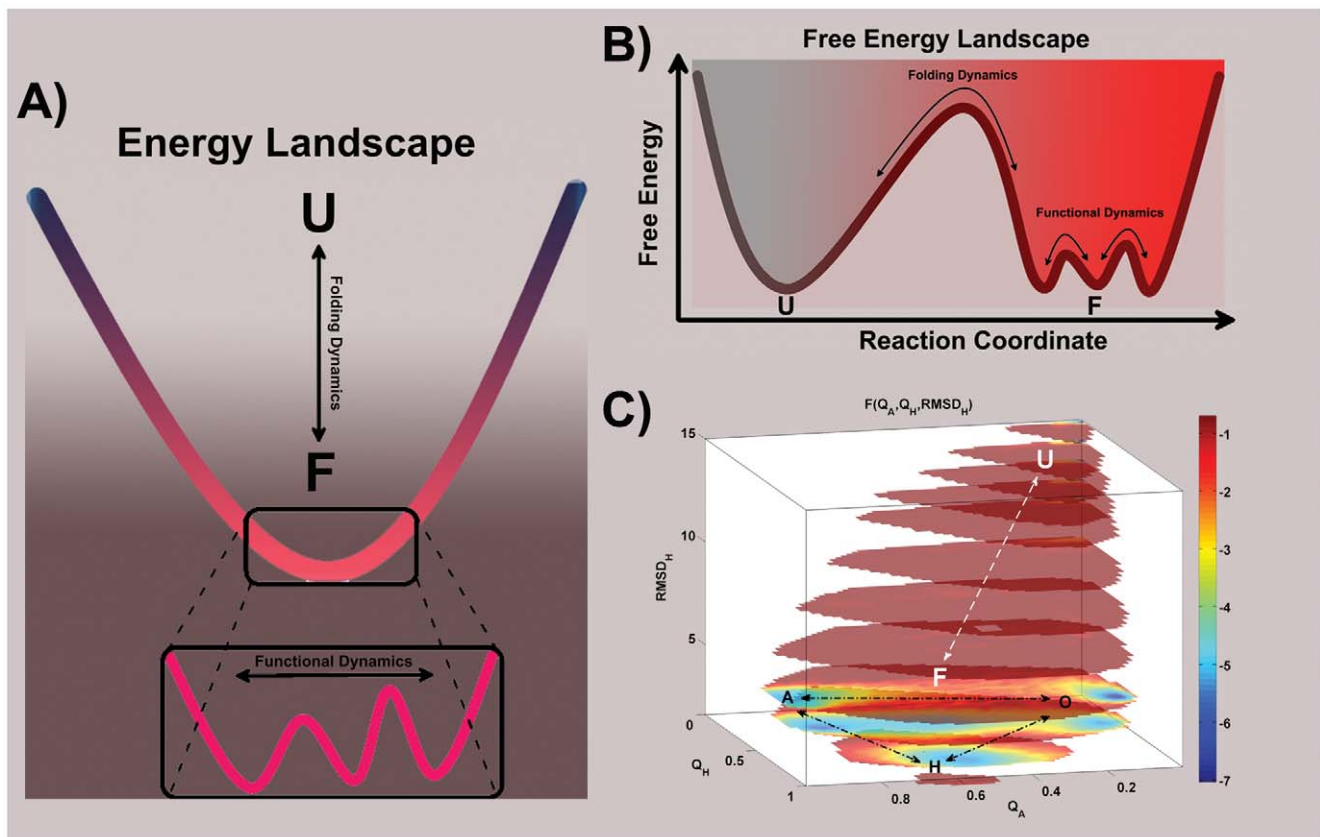


Figure 8. Schematic and actual representations of the folding landscape and functional landscape. A schematic energy landscape and a schematic free energy landscape are shown in (A) and (B), respectively. The unfolded state and folded states are denoted as U and F, respectively. (C) The actual free energy landscapes are shown by the high-dimensional free energy profile as a function of Q_A , Q_H , and $RMSD_H$ (in unit of Å) under modest ligand concentrations ($\epsilon_3 = 0.4$). The dashed arrows represent functional transitions between the open, apo-closed and holo-closed states. The $O \rightarrow A$ route that is hidden in two-dimensional free energy profiles can be visible in the four-dimensional free energy profiles. For better view of the functional landscape, the conformational regions with $RMSD_H$ beyond 15 Å are not shown.
doi:10.1371/journal.pcbi.1002471.g008

contributing to the roughness of energy landscape. There are several limitations in our models. The water solvent molecules are not explicitly considered here. Furthermore, the ligand is also implicitly introduced by ligand-mediated contacts. The implicit modeling of the ligands has an advantage in simplification of simulation and interpretation of the results, however it is unable to precisely account for the local nature of the interactions between ligand and protein. For example, the beta-cyclodextrin does not elicit domain closure for MBP, except for interactions with only the CTD [65]. In our simulation, we introduced the ligand-mediated interactions between NTD and CTD to reflect the ligand binding. So, our simulation is only able to represent the binding effect for these physiological ligands (such as maltose, maltotriose) with the exception of beta-cyclodextrin which interacts with CTD only. This important feature is also missing in the other models [5–10,12–15]. Indeed, we have further refined this model by introduction of explicit ligands. The improved model has been applied in this protein and another allosteric system and the related works will be submitted to be published soon.

Taken together, we propose that the conformational dynamics of MBP can be unraveled by assuming a triple-basin energy landscape (whose depths are modulated by ligand) that corresponds to distinct but related topological states. This work demonstrates that a multiple native basin biased landscape which follows the principle of

minimal frustration is sufficient to fold and function with complex topologies. From this perspective, we extend the application of the funneled energy landscape that explains how most proteins fold efficiently and robustly to their structures in functional transitions. Furthermore, we expect that our model can be extended to address more complex allosteric systems.

Materials and Methods

Depending on how pairwise interactions between residues are treated, current multiple-basin structure-based models can be broadly divided into two categories: microscopic and macroscopic. In macroscopic models, techniques borrowed from the Marcus theory of electron transfer [66,67] or an exponential Boltzmann-weighting method [12] are used to construct a double-well potential with two smoothly linked energy basins. In these models, pairwise favorable interactions in each reference structure are biased to its respective native basin. In contrast, with microscopic models [4,5,13], the two reference energy surfaces are integrated into two-well potentials for each pair of specific interactions. This results in a rougher landscape with more possible macroscopic states other than native basins. In addition, there is another method [6] that belongs to the type of models between microscopic and macroscopic level. In this approach, a potential based on one reference structure (ligand-free state) was modified

by the addition of a bias favoring interactions present in another reference structure (ligand-bound state). The energy potential in the model is constructed by combining multiple single-well potentials without redundant interactions. In other words, it mixes the contact maps rather than modifying the form of potential function. In the present work, we developed a coarse-grained model as an extension of such a model.

Triple-well model

To account for the side-chain dynamics, especially for active site, we developed a mixed coarse-grained model in which part of amino acids are represented by two beads. The coarse-graining process is similar to SMOG tool [68]. Based on the fact that the interface between the NTD and CTD is rich in charged residues, we introduced the electrostatic interactions into the model. In addition, we introduced a pseudo-ligand into the bound simulation by adding selected ligand-mediated interactions to the potential, an approach that has been used before in the double-well coarse-grained simulation [6,11,40]. The total Hamiltonian for the MBP system is given by the expression:

$$U_{total}(\Gamma_O, \Gamma_A, \Gamma_H) = U_{backbone} + U_{attraction} + U_{repulsive} + U_{charge}$$

The total energy is divided into backbone, non-bonded and electrostatic interactions. The backbone interaction $U_{backbone}$ maintains the geometry and local bias. The non-bonded interaction can be partitioned into two components, an attraction term $U_{attraction}$ to provide the triple-basin bias by a mixed contact map and a repulsive term $U_{repulsive}$ to provide the excluded volume.

For the last term U_{charge} , we used the Debye-Huckel potential to introduce the electrostatic interactions [69]. Water and ions were incorporated implicitly into the interaction model as the dielectric constant and the Debye screening length. We have tested the DH model using an extensive set of dielectric constants ranging between 40 and 160 and ion concentrations ranging from 0 to 0.20 M. The results of DH model with different parameters are summarized in Figure S12 in *Text S1*. Changing the dielectric constant or the salt concentration did not significantly affect the functional dynamics of MBP. Finally, we employed a salt concentration 0.10 M and a dielectric constant 80. All the analysis in manuscript was based on this values except where specified. Of course, the DH model is valid for low salt concentration and mainly for dilute solution. See *Text S1* for the detailed descriptions of these terms and the accompanying parameters of the force field.

References

- Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 103: 11844–11849.
- Bahar I, Lezon TR, Yang LW, Eyal E (2010) Global dynamics of proteins: Bridging between structure and function. *Annu Rev Biophys* 39: 23–42.
- Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 35: 539–546.
- Chu JW, Voth GA (2007) Coarse-grained free energy functions for studying protein conformational changes: A double-well network model. *Biophys J* 93: 3860–3871.
- Lu Q, Wang J (2008) Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc* 130: 4772–4783.
- Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational transitions of adenylate kinase: Switching by cracking. *J Mol Biol* 366: 1661–1671.
- Whitford PC, Gosavi S, Onuchic JN (2008) Conformational transitions in adenylate kinase. allosteric communication reduces misligation. *J Biol Chem* 283: 2042–8.
- Arora K, Brooks CL (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* 104: 18496–18501.
- Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci U S A* 101: 511–516.
- Andrews BT, Gosavi S, Finke JM, Onuchic JN, Jennings PA (2008) The dual-basin landscape in gfp folding. *Proc Natl Acad Sci U S A* 105: 12283–12288.
- Okazaki KI, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* 105: 11182–11187.
- Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure* 13: 1755–1763.
- Zuckerman DM (2004) Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J Phys Chem B* 108: 5127–5137.
- Zhang BW, Jasnow D, Zuckerman DM (2007) Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc Natl Acad Sci U S A* 104: 18043–18048.
- Yang S, Roux B (2008) Src kinase conformational activation: thermodynamics, pathways, and mechanisms. *Plos Comput Biol* 4: e1000047.

Simulation protocols

Simulations were performed with Gromacs 4.0.5 [70]. Reduced units were used for all calculations. A time step of 0.0005 time units (or ps) was used and the simulation was coupled to a temperature bath via Langevin dynamics with a coupling time of 1.0. For each individual trajectory, the total simulation time was 10^6 time units. To ensure that the simulation is converged and the statistical errors are small enough, we simultaneously ran several independent simulations using the same parameters set. Simulations were performed at $T=0.5$, and this temperature is used throughout the article, except where specified. Systems were initialized arbitrarily in one of the three native states.

Reaction coordinates

Q (and also RMSD) has been suggested to be useful when studying the folding landscape. Unfortunately, for functional dynamics which occurs at the bottom of the energy funnel, the traditional folding Q is unable to monitor the process due to the fact that most of the contacts shared by the native states are not needed to be broken. Fortunately, the reaction coordinates from protein folding can be borrowed to describe the functional landscape with minor modification. That is, to measure the formation of state-specific contacts instead of all native contacts. We can denote this quantity as the state-specific Q fraction or functional Q. The state-specific Q fraction can distinguish the conformational transitions between the native basins. This reaction coordinate has also been successfully used in other allosteric studies [18,40]. See more details in *Text S1*.

Supporting Information

Text S1 Supporting information of multi-state functional dynamics of MBP. (PDF)

Acknowledgments

We thank Peter G Wolynes for careful reading of the manuscript. Y.W. would like to thank Feng Zhang for help to plot the triple-basin landscape and Xiakun Chu for insightful discussions about the simulation details. We acknowledge the High Performance Computing Center (HPCC) of Jilin University for supercomputer time.

Author Contributions

Conceived and designed the experiments: EW JW. Performed the experiments: YW. Analyzed the data: YW CT. Contributed reagents/materials/analysis tools: EW JW. Wrote the paper: YW CT JW.

16. Itoh K, Sasai M (2010) Entropic mechanism of large actuation in allosteric transition. *Proc Natl Acad Sci U S A* 107: 7775–7780.
17. Hyeon C, Jennings PA, Adams JA, Onuchic JN (2009) Ligand-induced global transitions in the catalytic domain of protein kinase a. *Proc Natl Acad Sci U S A* 106: 3023–3028.
18. Yao XQ, Kenzaki H, Murakami S, Takada S (2010) Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nat Commun* 1: 117.
19. Kenzaki H, Koga N, Hori N, Kanada R, Li W, et al. (2011) Cafemol: A coarse-grained biomolecular simulator for simulating proteins at work. *J Chem Theor Comput* 7: 1979–1989.
20. Dwyer MA, Hellinga HW (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* 14: 495–504.
21. Medintz IL, Deschamps JR (2006) Maltose-binding protein: a versatile platform for prototyping biosensing. *Curr Opin in Biotechnol* 17: 17–27.
22. Duplay P, Szmelcman S, Bedouelle H, Hofnung M (1987) Silent and functional-changes in the periplasmic maltose-binding protein of *escherichia-coli-k12* .1. transport of maltose. *J Mol Biol* 194: 663–673.
23. Duplay P, Szmelcman S (1987) Silent and functional-changes in the periplasmic maltose-binding protein of *escherichia-coli-k12* .2. chemotaxis towards maltose. *J Mol Biol* 194: 675–678.
24. Spurlino JC, Lu GY, Quijcho FA (1991) The 2.3-Å resolution structure of the maltose-binding or maltodextrin-binding protein, a primary receptor of bacterial active-transport and chemotaxis. *J Biol Chem* 266: 5202–5219.
25. Sharff AJ, Rodseth LE, Spurlino JC, Quijcho FA (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the 2 domains of the maltodextrin binding-protein involved in active-transport and chemotaxis. *Biochemistry* 31: 10657–10663.
26. Quijcho FA, Spurlino JC, Rodseth LE (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport chemosensory receptor. *Structure* 5: 997–1015.
27. Duan XQ, Hall JA, Nikaido H, Quijcho FA (2001) Crystal structures of the maltodextrin/maltose-binding protein complexed with reduced oligosaccharides: Flexibility of tertiary structure and ligand binding. *J Mol Biol* 306: 1115–1126.
28. Skrynnikov NR, Goto NK, Yang DW, Choy WY, Tolman JR, et al. (2000) Orienting domains in proteins using dipolar couplings measured by liquid-state nmr: Differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J Mol Biol* 295: 1265–1273.
29. Evenas J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R, et al. (2001) Ligand-induced structural changes to maltodextrin-binding protein as studied by solution nmr spectroscopy. *J Mol Biol* 309: 961–974.
30. Millet O, Hudson RP, Kay LE (2003) The energetic cost of domain reorientation in maltose-binding protein as studied by nmr and uorescence spectroscopy. *Proc Natl Acad Sci U S A* 100: 12700–12705.
31. Aggarwal V, S RK, Rajaram S, MM B, Varadarajan R, et al. (2011) Single-molecule studies of the parallel unfolding pathways of maltose binding protein (mbp). *Biophys J* 100: 481a–481a.
32. Marvin JS, Hellinga HW (2001) Manipulation of ligand binding affinity by exploitation of conformational coupling. *Nat Struct Biol* 8: 795–798.
33. Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic nmr. *Nature* 449: 1078–U12.
34. Zhuravlev PI, Papoian GA (2010) Protein functional landscapes, dynamics, allostery : a tortuous path towards a universal theoretical framework. *Q Rev Biophys* 43: 295–332.
35. Stockner T, Vogel H, Tieleman D (2005) A salt-bridge motif involved in ligand binding and large-scale domain motions of the maltose-binding protein. *Biophys J* 89: 3362–3371.
36. Bucher D, Grant BJ, Markwick PR, McCammon JA (2011) Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS Comput Biol* 7: e1002034.
37. Bucher D, Grant BJ, McCammon JA (2011) Induced fit or conformational selection? the role of the semi-closed state in the maltose binding protein. *Biochem* 50: 10530–9.
38. Kondo HX, Okimoto N, Morimoto G, Taiji M (2011) Free-energy landscapes of protein domain movements upon ligand binding. *J Phys Chem B* 115: 7629–36.
39. Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* 120: 11919–11929.
40. Daily MD, Phillips GN, Cui QA (2010) Many local motions cooperate to produce the adenylate kinase conformational transition. *J Mol Biol* 400: 618–631.
41. Wang J, Wang Y, Chu X, Hagen SJ, Han W, et al. (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor ia3 to its target enzyme. *PLoS Comput Biol* 7: e1001118.
42. Li W, Wolynes PG, Takada S (2011) Frustration, specific sequence dependence, and nonlinearity in large-amplitude actuations of allosteric proteins. *Proc Natl Acad Sci U S A* 108: 3504–9.
43. Kurakin A (2010) Order without design. *Theor Biol Med Model* 7: 12.
44. Staii C, Wood DW, Scoles G (2008) Ligand-induced structural changes in maltose binding proteins measured by atomic force microscopy. *Nano Lett* 8: 2503–2509.
45. Hammes GG, Chang YC, Oas TG (2009) Conformational selection or induced fit: A ux description of reaction mechanism. *Proc Natl Acad Sci U S A* 106: 13737–13741.
46. Zhou HX (2010) From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J* 98: L15–L17.
47. Marinelli F, Pietrucci F, Laio A, Piana S (2009) A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *Plos Comput Biol* 5: e1000452.
48. Palmer r AG (2004) Nmr characterization of the dynamics of biomacromolecules. *Chem Rev* 104: 3623–40.
49. Miller DM, Olson JS, Pugrath JW, Quijcho FA (1983) Rates of ligand-binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J Biol Chem* 258: 3665–3672.
50. Gosavi S, Chavez LL, Jennings PA, Onuchic JN (2006) Topological frustration and the folding of interleukin-1 beta. *J Mol Biol* 357: 986–996.
51. Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci U S A* 105: 14844–8.
52. Sulkowska JI, Sulkowski P, Szymczak P, Cieplak M (2008) Stabilizing effect of knots on proteins. *Proc Natl Acad Sci U S A* 105: 19714–9.
53. Novokhatny V, Ingham K (1997) Thermodynamics of maltose binding protein unfolding. *Protein Sci* 6: 141–146.
54. Cunco MJ, Changela A, Beese LS, Hellinga HW (2009) Structural adaptations that modulate monosaccharide, disaccharide, and trisaccharide specificities in periplasmic maltose-binding proteins. *J Mol Biol* 389: 157–166.
55. Lai ZZ, Lu Q, Wang J (2011) Exploring the thermodynamic landscape, kinetics, and structural evolution of a protein conformational transition with a microscopic double-well model. *J Phys Chem B* 115: 4147–59.
56. Wlodarski T, Zagrovic B (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A* 106: 19346–51.
57. Silva DA, Bowman GR, Sosa-Peinado A, Huang X (2011) A role for both conformational selection and induced fit in ligand binding by the lao protein. *Plos Comput Biol* 7: e1002054.
58. Vertessy BG, Orosz F (2011) From “actuation fit” to “conformational selection”: evolution, rediscovery, and integration of a concept. *Bioessays* 33: 30–4.
59. Felder CB, Graul RC, Lee AY, Merkle HP, Sadee W (1999) The venus ytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. *AAPS PharmSci* 1: E2.
60. Mao B, Pear MR, McCammon JA, Quijcho FA (1982) Hinge-bending in l-arabinose-binding protein. the “venus’s-ytrap” model. *J Biol Chem* 257: 1131–3.
61. Bermejo GA, Strub MP, Ho C, Tjandra N (2010) Ligand-free open-closed transitions of periplasmic binding proteins: the case of glutamine-binding protein. *Biochemistry* 49: 1893–902.
62. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *J Mol Biol* 298: 937–953.
63. Schug A, Onuchic JN (2010) From protein folding to protein function and biomolecular binding by energy landscape theory. *Curr Opin Pharmacol* 10: 709–714.
64. Perilla JR, Beckstein O, Denning EJ, Woolf TB (2011) Computing ensembles of transitions from stable states: Dynamic importance sampling. *J Comp Chem* 32: 196–209.
65. Sharff AJ, Rodseth LE, Quijcho FA (1993) Refined 1.8-Å structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein. *Biochemistry* 32: 10553–9.
66. Marcus RA, Sutin N (1985) Electron transfers in chemistry and biology. *Biochim Biophys Acta* 811: 265–322.
67. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, protein-quakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A* 100: 12570–12575.
68. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) Smog@ctbp: simplified deployment of structure-based models in gromacs. *Nucleic Acids Res* 38: W657–61.
69. Azia A, Levy Y (2009) Nonnative electrostatic interactions can modulate protein folding: Molecular dynamics with a grain of salt. *J Mol Biol* 393: 527–542.
70. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theor Comput* 4: 435–447.

SI Text

Yong Wang ^{*}, Chun Tang [†], Erkwang Wang ^{*,‡}, and Jin Wang ^{*,§¶}

^{*}State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, 130022, P.R. China, [†]State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan, Hubei 430071, China, and [§]Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, USA

Model and Methods

Energy Hamiltonian. To extend a structure-based model to systems with multiple basins, we modified the contact map and the dihedral angles. Note that all the residue indexes are rearranged from 1 to 366 in the model. To include the side chain dynamics on the conformational change of enzyme, we developed a coarse-grained model in which each amino acid is represented by one or two beads dependent of its location. One bead (named C_A in our model), representing the backbone, is located on the C_α atom, and one bead (name C_B), representing the side chain, is located on the center of mass or the farthest heavy atom of the side chain, depending on its residue characterize (see details in section of *Side-chain Interactions*). Based on the fact that the interface between the NTD and CTD is lined by negatively charged residues, that are responsible for an array of hydrogen bonds with the sugar substrate in holo-close form of MBP, we introduced the electrostatic interactions into the two-bead model to study the effect of charged residues on the conformational change of MBP. The two-bead structure-based hamiltonian is given by the expression:

$$\begin{aligned} U_{total}(\Gamma_O, \Gamma_A, \Gamma_H) &= U_{backbone} + U_{attraction} + U_{repulsive} + U_{charge} \\ &= \sum_{bonds} K_r(r - r_0)^2 \\ &\quad + \epsilon_p * [U_{angle}(\gamma^{non-hinge}) + \epsilon_{hinge} * U_{angle}(\gamma^{hinge})] \\ &\quad + \epsilon_1 * U_{LJ}(\gamma^C) \\ &\quad + \epsilon_2 * [\epsilon_o * U_{LJ}(\gamma^O) + \epsilon_a * U_{LJ}(\gamma^A) + \epsilon_h * U_{LJ}(\gamma^H)] \\ &\quad + \epsilon_3 * U_{LJ}(\gamma^{ligand}) \\ &\quad + \sum_{non-native} K_{repulsion} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \\ &\quad + K_{coulomb} * B(k) \sum_{i,j} \frac{q_i * q_j * exp(-kr_{ij})}{\epsilon_r r_{ij}} \end{aligned}$$

The total energy is divided into backbone, non-bonded and electrostatic interactions. The backbone interaction $U_{backbone}$ maintains the geometry and local bias. The non-bonded interaction can be partitioned into two components: attraction term $U_{attraction}$ is contained within a contact map to provide the tertiary bias; and repulsive term $U_{repulsive}$ is used to provide the excluded volume. The contacts are determined from the given PDB structures. Any atoms not interacting through a contact, bond, angle or dihedral, are considered 'non-contacts' and interact only through excluded volume. The key to construct a triple-basin model is to define a mixed contact map which integrates multiple structural information together. The U_{charge} term introduces the electrostatic interactions. A detailed description of a structure-based model can be found elsewhere (1).

In the MD simulation, we used coordinates of PDB 1OMP, 2V93 and 3MBP to represent the open, apo-close and holo-close state, respectively. A triple-well potential was modelled according to the three PDB coordinates. Note that, we use the first MODEL of 2V93 to import the information of the apo-close state.

Repulsive interaction.

$$U_{repulsive} = \sum_{non-native} K_{repulsion} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12}$$

σ_{NC} is the excluded distance between non-native pairs to provide excluded volume repulsion. The repulsive radii is 4.0 and 2.5 Å for C_A and C_B , respectively. r_{ij} is the distance between bead i and bead j. The strength of the repulsive term $K_{repulsion}$ is 1 kJ/mol. All possible contact pairs existing in contact map of each state are not considered in the repulsive term.

Construction of triple-basin potential.

$$\begin{aligned} U_{attraction} &= \epsilon_1 * U_{LJ}(\gamma^C) + \epsilon_2 * [\epsilon_o * U_{LJ}(\gamma^O) + \\ &\quad \epsilon_a * U_{LJ}(\gamma^A) + \epsilon_h * U_{LJ}(\gamma^H)] + \epsilon_3 * U_{LJ}(\gamma^{ligand}) \\ U_{LJ}(\gamma^X) &= \sum_{\gamma^X} [5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10}] \end{aligned}$$

The non-bonded interactions are defined by contact map obtained with CSU (2). CSU acknowledges a contact between two atoms based on the intersection of their putative surfaces and their hydrophobicity. Each contact is regarded as an attractive interactions between the two involved amino acids. In our two-bead model, these interactions are divided into different contributions. The possible combinations are contacts of the type $C_A - C_A$, $C_A - C_B$ and $C_B - C_B$. In our model, we only include $C_A - C_A$ and $C_B - C_B$ contacts, although $C_A - C_B$ pairs still are included in repulsive term given by $\sigma_{ij} = \sqrt{4.0 * 2.5} = 3.16 \text{Å}$. Including the $C_A - C_B$ contacts into the model should not qualitatively change the folding mechanism (3).

The concrete details of building the mixed contact map are as follow:

a) **Total contact library γ^{All} .** We build a contact library by collecting all the $C_A - C_A$ and $C_B - C_B$ pairs in the contact maps obtained from open, apo-close and holo-close PDB coordinates. There are 1359 contacts (1313 $C_A - C_A$ pairs and 46 $C_B - C_B$ pairs) in the total library γ^{All} .

b) **Reference distances.** We calculate the distances r_{ij}^O , r_{ij}^A and r_{ij}^H between the two beads i and j in the library. r_{ij}^O , r_{ij}^A and r_{ij}^H represent the beads' distance in open, apo-close and holo-close states, respectively. The often encountered problem is these contacts appear in the three states with different

[‡]E-mail:ekwang@ciac.jl.cn

[¶]To whom correspondence should be addressed. E-mail: jin.wang.1@stonybrook.edu

r_{ij} . Thus, care must be taken while constructing a mixed attractive potential with multiple basins.

c) **Core sublibrary γ^C .** We define $R_{ij}(X/Y) = \frac{r_{ij}^X}{r_{ij}^Y}$, X and Y can be O, A and H. For a bead pair i and j , if any $R_{ij}(X/Y) < R_{cut}$, this bead pair is considered as a member of core contact library (γ^C). We set $R_{cut} = 1.1$ for $C_A - C_A$ pairs and $R_{cut} = 1.25$ for $C_B - C_B$ pairs. There are 980 $C_A - C_A$ pairs and 14 $C_B - C_B$ pairs in core contact library γ^C .

d) **State-specific sublibraries γ^O , γ^A and γ^H .** If any possible $R_{ij}(X/Y) > 1.5$ is in a pair, this pair will be considered as a member of unique contact library which contains three subsets, open library γ^O , apo-close library γ^A and holo-close library γ^H . Note that any contact libraries contain three parameters, atom indexes of two beads and corresponding optimal distance. Any other conflicting contacts which may strain the system are disregarded. There are 31 contacts (all is $C_A - C_A$ pairs) in γ^O . There are 55 contacts (47 $C_A - C_A$ pairs and 8 $C_B - C_B$ pairs) in γ^H . There are 40 contacts (23 $C_A - C_A$ pairs and 17 $C_B - C_B$ pairs) in γ^A .

In summary, γ^O , γ^A and γ^H combine the structural information of open, apo-close and holo-close state, furthermore the additional contact library γ^{ligand} are added to represent ligand binding. The strengths of the core contacts and the state-specific contacts are rescaled by ϵ_1 and ϵ_2 , respectively. And the strength of ligand binding contacts is rescaled by ϵ_3 . Furthermore, ϵ_o , ϵ_a and ϵ_h are introduced to modulate the relatively stability of O, A and H basins, respectively.

Side-chain Interactions. A comparison between the apo-close (PDB code 2v93) and holo-close (PDB code 3mbp) forms of MBP suggests that the side chains in active site groove play important roles in binding with maltose ligands. A C_α -backbone model with uniform interactions is lack of resolution to describe this feature. Thus, a number of important features associated with the conformational transition of MBP are involved with side-chain dynamics. To introduce side chains into a coarse-grained structural based model, we have modified the C_α model. Because many side-chain interactions in active site of MBP are involved between charges residues (Asp, Glu, Lys, Arg), we use the most distant heavy atom (N,C,O) from the C_α to model the charged residues. For the non-charged residues, the interaction centers of side chains are modeled by using center of mass of the side-chain atoms in the native structures. We did not include the side chains, except those around the active site, so that in the model to minimize the computational cost. Including other side chains into the model should not qualitatively change the dynamics.

Ligand-mediated Contacts (γ^{ligand}). We introduce a pseudo-ligand into the bound simulation by adding selected ligand-mediated interactions to the structured based potential, an approach that has been used before in the double-well coarse-grained simulation (4, 5). An explicit coarse-grained ligand representation was previously used for simulation of ligand-induced global transitions in the catalytic domain of protein kinase A (6). However, such a representation will make the dynamics more complex.

i) We select the residues around ligand in the active site. It includes D10, K11, E40, W58, E149, P150, Y151, W226, M326, W336, Y337, R340 (Fig. S1).

ii) We calculate the distance r_{ij} between the C_α atoms of any

two ligand-binding residues i and j (in the above list).

iii) We exclude local contacts ($|i - j| < 4$ residues) and intra-domain contacts.

vi) The ligand-mediated contacts library γ^{ligand} is constructed by the contacts in the holo-close coordinate with $C_A - C_A$ distances that are over 40% further apart ($R_{ij}(O/H)$ or $R_{ij}(A/H) > 1.4$) while in the open form or apo-close form. Finally, we identified 14 ligand-mediated interactions based on the maltotriose-binding close structure (3MBP).

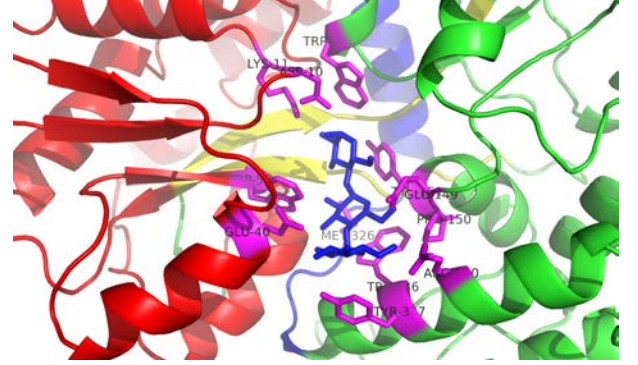


Fig. 1. Important residues around active site. Ligand (maltotriose here) is represented by blue sticks.

Backbone Potential and the Flexibility of Hinge Regions.

$$\begin{aligned}
 U_{backbone} &= U_{bond} + \epsilon_p * [U_{angle}(\gamma^{non-hinge}) + \epsilon_{hinge} * U_{angle}(\gamma^{hinge})] \\
 U_{bond} &= \sum_{bonds} K_r(r - r_0)^2 \\
 U_{angle}(\Gamma_X) &= U_{CX-CX-CX}(\Gamma_X) + U_{CX-CX-CX-CX}(\Gamma_X) \\
 U_{CX-CX-CX}(\Gamma_X) &= \sum_{angles} K_\theta(\theta - \theta_0)^2 \\
 U_{CX-CX-CX-CX}(\Gamma_X) &= \sum_{dihedrals} K_\phi[(1 - \cos(\phi - \phi_0)) + 0.5(1 - \cos(3(\phi - \phi_0)))]
 \end{aligned}$$

The total backbone energy is divided into bond stretching, angle bending, torsion interactions. The bonded energies U_{bond} are summed over the energy of all co-valent bonds. $K_r = 10000$ kJ/(mol nm²) is the bond constant, r is the distance between the two bonded atoms and r_0 is the reference distance of these atoms in the native structure. The angular energy U_{angle} has the angle constant $K_\theta = 20$ kJ/mol. θ is the angle between two adjacent bonds, and θ_0 is the reference angle in the native structure.

The dihedral energy $U_{dihedral}$ possesses the dihedral constant K_ϕ and the angle ϕ between the two planes formed by four connected atoms. We distinguish between different dihedral angles, dependent on the type of the involved atoms. If all four atoms are C_A , $K_\phi = 0.344$ kJ/mol. For $CB-CA-CA-CB$ dihedral angles and $CA-CA-CA-CB$ dihedral angles, $K_\phi = 0.086$ kJ/mol. Improper torsion angles (or improper dihedral angles) are implemented using the same equation. Improper dihedral angles have a maximum value at the cis conformation and help to maintain the system's chi

rality.

The hinge regions were computationally identified from differences in both pseudo-angles and pseudo-dihedral angles between any two native conformations of open, apo-close and holo-close states. Hinges are located wherever pseudo angle or pseudo dihedral angle differences are greater than threshold values. In this article, the threshold values for angle bending term $U_{CX-CX-CX}(hinge)$ is 15 degree and for dihedral angle term $U_{CX-CX-CX-CX}(hinge)$ is 40 degree corresponding to $1KJ/mol$. The list of hinge regions is shown in Fig. S2. Note that we only list the backbone angles (consecutive CA-CA) and dihedral angles (consecutive CA-CA-CA-CA) in table.

Pseudo-angles Table in Hinge Regions									
i-1	i	i+1	O	A	H	O-A	A-H	O-H	
250	251	252	109.11	120.02	103.6	10.91	16.42	5.51	
253	254	255	111.89	132.06	101.58	20.17	30.48	10.31	
323	324	325	119.08	113.24	130.78	5.84	17.54	11.7	
327	328	329	95.376	111.52	94.328	16.144	17.192	1.048	
Pseudo-dihedral angle Table in Hinge Regions									
i-1	i	i+1	i+2	O	A	H	O-A	A-H	O-H
252	253	254	255	256.96	292.03	245.48	35.07	46.55	11.48
309	310	311	312	210.39	176.14	217.57	34.25	41.43	7.18
322	323	324	325	264.17	55.12	250.82	150.95	164.3	13.35
327	328	329	330	109.68	93.39	134.05	16.29	40.66	24.37

Fig. 2. Tables of hinge regions determined according to our criterion. Only the backbone angles (consecutive CA-CA-CA) and dihedral angles (consecutive CA-CA-CA-CA) are listed in table.

Electrostatic Interaction.

$$U_{charge} = K_{coulomb} * B(k) \sum_{i,j} \frac{q_i * q_j * exp(-kr_{ij})}{\epsilon_r r_{ij}} \quad [1]$$

The electrostatic potential was represented by the Debye-Huckel model, which is a linearization of the Poisson-Boltzmann equation. All charged residues were given a full charge according to their electrostatic charge at neutral pH. The charge was placed on the C_B bead. The Debye-Huckel theory predicts the range of electrostatic influences of an ion to be the Debye screening length $1/k$. Linearization of the Poisson-Boltzmann equation yields the following relation $k^2 = \frac{8\pi N_A e^2 \rho_A}{1000 \epsilon k_B T}$, where N_A is Avogadro's number, ρ_A is the solvent density, e is the proton charge, ϵ is the solvent dielectric constant, and C_s is the ionic concentration in molar unit. For monovalent salt at room temperature and with $\epsilon = 80$, $k = 0.32\sqrt{C_s} \text{Å}^{-1}$. q_i is the point charge of the i th bead, r_{ij} is the distance between two charged beads, $K_{coulomb} = 4\pi\epsilon_0 = 332kcalmol^{-1} = 138kJmol^{-1}nm^{-1}$.

B_k is the salt-dependent coefficient, and r_{ij} is the distance between charged beads i and j . For the calculation of the screening factor k and the salt-dependent coefficient $B(k)$, solvent density was taken as $1kg/L$ and the ion radius was taken as 1.4Å . More details of the Debye-Huckel model can be found elsewhere (7, 8).

In this study, we set $\epsilon_r = 80$, the dielectric constant of water is at near-ambient conditions. The Debye screening length $1/k$ depends on the salt concentration. For simplicity, we focus on physiological salt concentrations of 100 nM , which lead to $1/k = 10 \text{Å}$. Residue charges correspond to pH 7, such that $q_i = +e$ for Lys and Arg, $-e$ for Asp and Glu, where e is the elementary charge.

Superior Angle Models. Based on the difference of backbone angles in all possible states of MBP, we developed three mod-

els: Φ_O (using native angles in open state), Φ_A (using native angles in apo-close state) and Φ_H (using native angles in holo-close state). We found that Φ_O and Φ_H models were able to reproduce the three native states with reasonable stability. It seems that the energy landscape constructed from Φ_A potential is not smooth enough to make the system mis-fold in local minima other than native basins. This result suggests that the apo-close structure as a relatively unstable conformation, is not appropriate to be used as the basic structure to model the basin dynamics.

Functional Φ values

The functional Φ values can be approximately calculated through the following equation:

$$\Phi_{F_i}^{RP} = \frac{\langle Q_i \rangle^{R \cap P} - \langle Q_i \rangle^R}{\langle Q_i \rangle^P - \langle Q_i \rangle^R}$$

where $\langle Q_i \rangle$ is the thermal mean value of the number of two-body interaction contacts for residue i over all the corresponding states, and the $R \cap P$ subscript represents the transition state from R to P state. R and P subscripts represent reactant state and product state which can be any two of the open, apo-close and holo-close states, respectively. Note that, for these residues in which the difference between $\langle Q_i \rangle^P$ and $\langle Q_i \rangle^R$ is less than a certain cutoff value (using 1.0 in this article), their functional Φ_{F_i} values are not calculated.

It is important to clarify the similarities and differences in the definition and interpretation of Φ_F -values and protein folding Φ -values. Protein folding Φ -values give a measure of the interaction strength of a particular residue existing in native state which is also present in the folding transition state ensemble. Similarly, functional Φ -values measure the interaction strength involving a particular residue in the transition state from R state to P state. Note that, the reactant and product states are dependent on the definition. Both folding Φ and functional Φ values give a quantitative detection of a residue's importance in the transition state from reactant to product conformation. The main difference is, in the definition of reactant state and product state, that folding Φ calculation uses the unfolded state as reactant state and folded native state as product state. Overall, $\Phi_i = 1$ implies that this particular residue is crucial for change from reactant to product state, or is close to the product conformation at the transition state. On the other hand, if $\Phi_i = 0$, then it means this particular residue is not important for change from reactant to product state or close to the reactant state at the transition state. It is noteworthy that the best targets for mutations to perturb the transition rates are those residues for which Φ_{F_i} close to either 0 or 1.

Due to the complexity of conformational switching between multiple basins, $\langle Q_i \rangle^P$ can be less than $\langle Q_i \rangle^R$ which indicates that contacts formed in residue i at R state are more than P state. And $\langle Q_i \rangle^{R \cap P}$ can be out of the range of $\langle Q_i \rangle^P$ and $\langle Q_i \rangle^R$, although this case does not occur frequently. It can cause Φ_{F_i} value to be negative or even larger than 1 which are not easily interpreted in traditional folding Φ value analysis. Therefore, the functional Φ_F -values should be handled carefully.

Local Cracking

In this article, local unfolding was measured by calculating the average deviation in backbone dihedral angle comprised of residue $i-1$, i , $i+1$, $i+2$:

$$C_i^{RP} = 0.5 * \frac{|\langle \Psi_i \rangle^{R \cap P} - \Psi_i^R| + |\langle \Psi_i \rangle^{R \cap P} - \Psi_i^P| - |\Psi_i^R - \Psi_i^P|}{|\Psi_i^R - \Psi_i^P|}$$

where Ψ_i is the mean value of the dihedral angle for residue i over all the corresponding state, and the $R \cap P$ subscript represents the transition state between R state and P state. It is noteworthy that the C_i^{RP} we calculated reflects the potential cracking regions which may not contain the known hinge residues, such as E107, V257 and A308, due to the large $|\Psi_i^R - \Psi_i^P|$. Furthermore, we don't consider the residues whose $|\Psi_i^R - \Psi_i^P|$ are less than a certain threshold value (20 degree in the article). If C_i^{RP} is larger than 0.1, a local cracking is considered to occur.

Reaction Coordinates

The state of MBP is quantitatively measured by the fraction of formation of certain state-specific contacts, the so-called state-specific Q fraction or functional Q. The state-specific Q fraction has been used in previous conformational change studies (4, 9) and suggested to be a good reaction coordinate to capture the process of conformational change between multiple states. Namely, to monitor the closeness to apo-close state, we used a set of residue pairs that are in contact in the A state, but not in contact in the O state or the H state. These A-specific contacts monitor the closeness to the A state.

$$Q = \sum_{Open, Apo, Close} \frac{1 - (R_{ij} - D_0)/R_{ij}^{nat})^n}{1 - ((R_{ij} - D_0)/R_{ij}^{nat})^m}$$

Where, $m=20$, $n=10$, R_{ij}^{nat} is native distance of a pair atom i and atom j . $D_0 = 0.42 * R_{ij}^{nat}$.

Static Structure Analysis

The previous studies (10, 11) on the conformational equilibria between open and holo-close states show that interactions on the interface opposite the ligand binding cleft, located at some distance from the ligand binding pocket, are responsible for modulating the stability of the open conformation. It has an effect on the ligand-binding affinity by a "conformational coupling" mechanism (10). This functional interface, called "balancing interface", was considered to play a role of "molecular switch" that triggers the conformational turnover.

The balancing interface includes two important segments. One is a loop region (residues Y167 to D173) which is modestly conserved in MBPs from various organisms. While MBP is in open state, this loop interacts with NTD (N-terminal domain) to form a number of contacts. These contacts are broken in partially and fully closed state, resulting in higher flexibility for the loop (12, 13). Another segment is the linker2 (residues D310 to P330) which plays important roles not only in bridging the two domains, but also as the base of the ligand-binding groove. Several site-specific chemical modifications in this region, such as replacement of I325 with larger bulky groups (10), mutations of residues M317 and Q321 to alanine (11), have large impacts in destabilizing the open conformation and increasing the ligand binding affinity by almost two

orders of magnitude. As such, in the present work, we term the loop as "balancing loop" and the linker2 as "balancing linker" for more clear description of the process in details (see schematic diagram in Figure 1 in main text).

By comparing the structures of open, apo-close and holo-close states, we found that balancing loop forms contacts with NTD in open state and is free in apo-close and holo-close states (Figure 1 in main text). On the other hand, from open state to holo-close state, balancing linker moves away from NTD (5.6Å translation for C_α in Q321). This results in the contacts between balancing linker and NTD decrease, and several residues (including I325, M317, Q321) become more solvent-exposed. In the apo-close state, balancing linker is closer to NTD following a rotation of linker1 and linker2. However, the interacting site in NTD is different with that in open state due to the movement of linker2.

References

1. Clementi, C, Nymeyer, H, & Onuchic, J. N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins *J Mol Biol* **298**, 937–953.
2. Sobolev, V, Sorokine, A, Prilusky, J, Abola, E. E, & Edelman, M. (1999) Automated analysis of interatomic contacts in proteins *Bioinformatics* **15**, 327–32.
3. Oliveira, L. C, Schug, A, & Onuchic, J. N. (2008) Geometrical features of the protein folding mechanism are a robust property of the energy landscape: a detailed investigation of several reduced models *J Phys Chem B* **112**, 6131–6.
4. Daily, M. D, Phillips, G. N, & Cui, Q. A. (2010) Many local motions cooperate to produce the adenylate kinase conformational transition *J Mol Biol* **400**, 618–631.
5. Whitford, P. C, Miyashita, O, Levy, Y, & Onuchic, J. N. (2007) Conformational transitions of adenylate kinase: Switching by cracking *J Mol Biol* **366**, 1661–1671.
6. Hyeon, C, Jennings, P. A, Adams, J. A, & Onuchic, J. N. (2009) Ligand-induced global transitions in the catalytic domain of protein kinase a *Proc Natl Acad Sci USA* **106**, 3023–3028.
7. Azia, A & Levy, Y. (2009) Nonnative electrostatic interactions can modulate protein folding: Molecular dynamics with a grain of salt *J Mol Biol* **393**, 527–542.
8. Givaty, O & Levy, Y. (2009) Protein sliding along dna: dynamics and structural characterization *J Mol Biol* **385**, 1087–97.
9. Yao, X. Q, Kenzaki, H, Murakami, S, & Takada, S. (2010) Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations *Nat Commun* **1**, 117.
10. Marvin, J. S & Hellinga, H. W. (2001) Manipulation of ligand binding affinity by exploitation of conformational coupling *Nat Struct Biol* **8**, 795–798.

11. Telmer, P. G & Shilton, B. H. (2003) Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants *J Biol Chem* **278**, 34555–34567.
12. Quioco, F. A, Spurlino, J. C, & Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport chemosensory receptor *Structure* **5**, 997–1015.
13. Tang, C, Schwieters, C. D, & Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic nmr *Nature* **449**, 1078–U12.

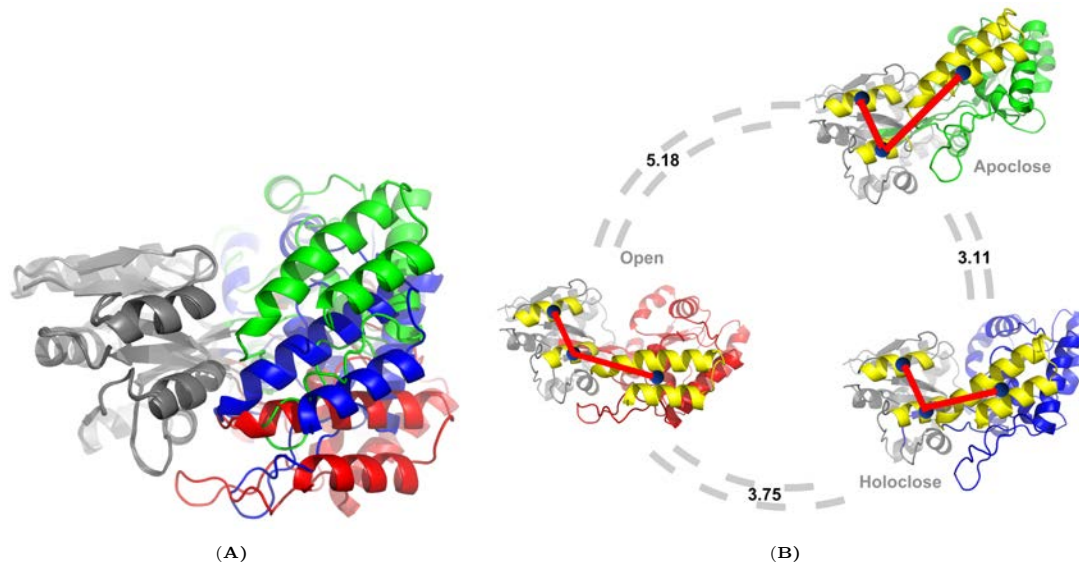


Fig. 3. Structure comparison (A) Structure comparison between open, apo-close and holo-close whose CTDs are colored in red, green and blue, respectively (NTD is grey). (B) Backbone RMSD between open and holo-close, between open and apo-close, and between apo-close and holo-close are 3.75, 5.18 and 3.11, respectively.

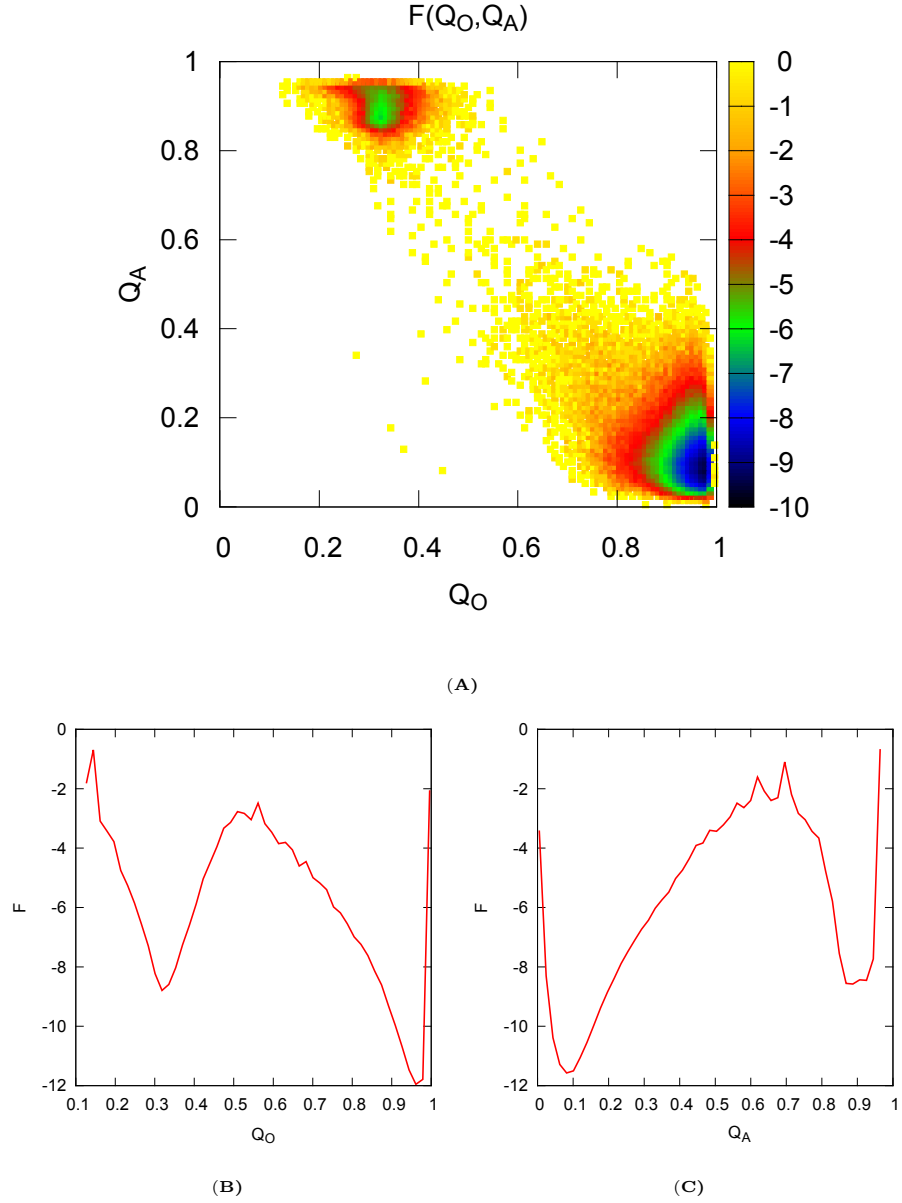


Fig. 4. Dynamical equilibrium between major open form and minor apo-close form of free MBP. These parameters are $\epsilon_1 = 1.0$, $\epsilon_2 = 0.4$, $\epsilon_o = 1.0$, $\epsilon_a = 1.0$, $\epsilon_h = 0.2$, $\epsilon_3 = 0.001$, $\epsilon_p = 0.5$, $\epsilon_{hinge} = 0.1$, with Φ_H potential. The population of open state is 93.96%. Apo-close state is 4.93%. Note that there is about 1.00% misfolded conformation. (A) The two-dimensional free energy surface is shown as a function of Q_O and Q_A . (B) The free energy profile is shown as a function of Q_O . (C) The free energy profile is shown as a function of Q_A . $(Q_A, Q_O) = (0.9, 0.3)$ and $(0.1, 0.9)$ for the apo-close minimum and open minimum, respectively. The free energy barrier from apo-close to open state is $6K_bT$, from open state to apo-close is $9K_bT$. In addition, transition state TS_{OA} between O and A basin is located at $(Q_A, Q_O) = (0.6, 0.5)$.

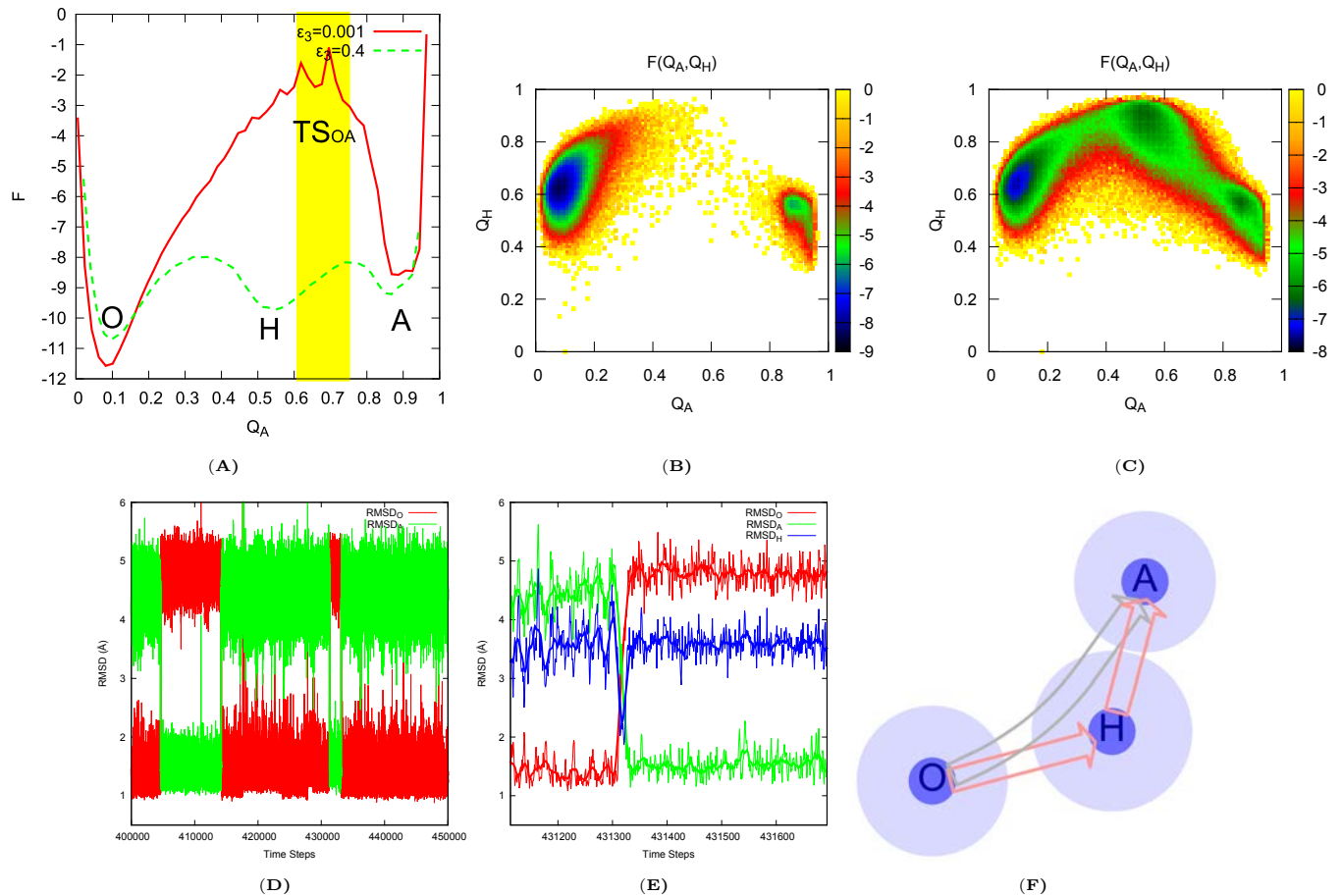


Fig. 5. (A) Free energy as a function of Q_A without ligand binding ($\epsilon_3 = 0.001$ by red solid line) and with modest ligand binding ($\epsilon_3 = 0.4$ by dashed green line). The region of transition state ensemble of O-A transition are colored in yellow. It is evident that transition state of O-A transition is not holo-close state. (B) Free energy profiles $F(Q_A, Q_H)$ of free MBP indicates O-A transition. (C) Free energy profiles $F(Q_A, Q_H)$ of MBP under modest ligand concentration indicates O-H and H-A transition. O-A transitions are hidden in the free energy profile. (D) Typical kinetic trajectory of O-A transitions. (E) One typical O-A transition. The smoother line is a running average of neighbor 20 samples. (F) Schematic basin transitions. It is difficult to distinguish the directly O-A transition (grey arrow) and indirectly O-A transition (pink arrows) passing H basin from free energy profiles.

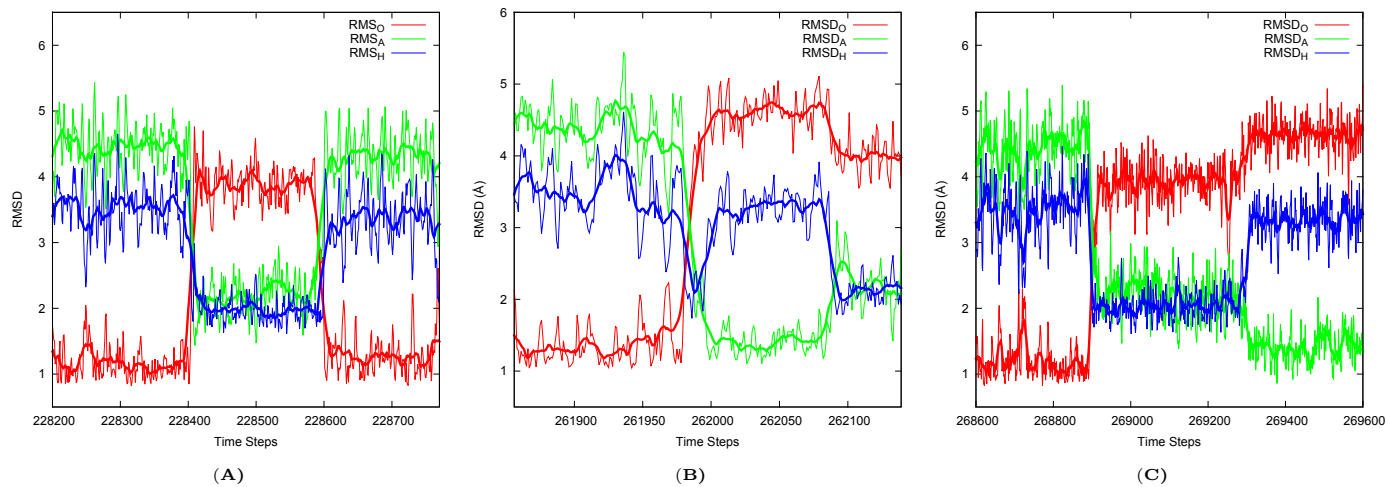


Fig. 6. Typical kinetic trajectories. (A) Induce fit route $O \rightarrow H$. (B) Population shift route $O \rightarrow A \rightarrow H$. The kinetic trajectory shows that a thermodynamically invisible path actually exist. (C) $O \rightarrow H \rightarrow A$ route.

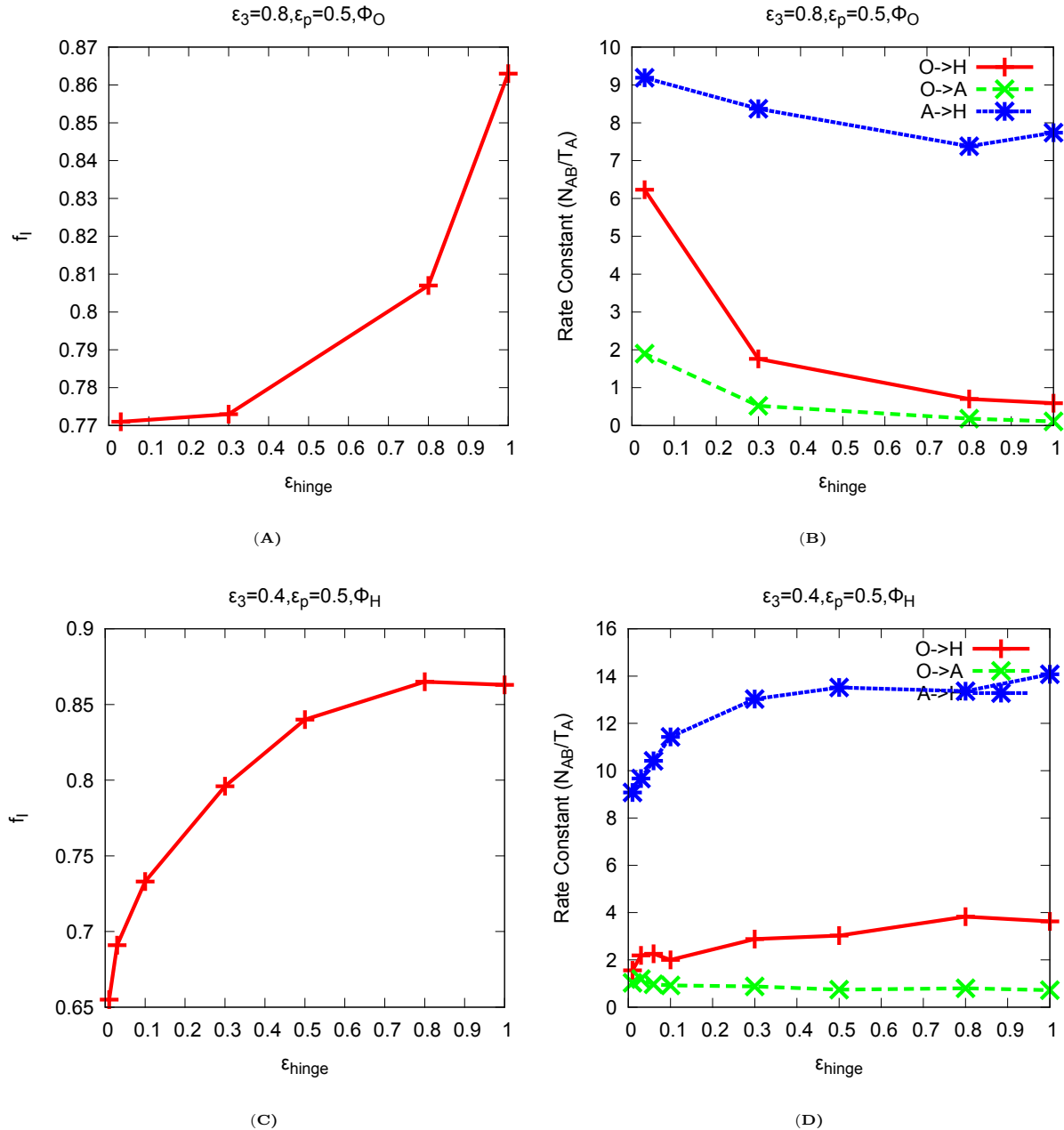


Fig. 7. Effects of hinge flexibility on the kinetics mechanism and transition rate constants. The relationships between fractional flux of IF routes f_I and hinge flexibility with Φ_O (a) and Φ_H models (c) are shown. Furthermore, the influences of hinge flexibility on transition rate constants are shown in (b) with Φ_O model and (d) with Φ_H model. Increasing hinge rigidity in both models can facilitate the conformational transition process of MBP along induced fit pathway. However, the influences of hinge flexibility on the transition rate have big differences. In Φ_O model hinge flexibility can greatly accelerate the O-H. However the corresponding fractional flux f_I decreases as hinge flexibility increases. This can be explained as the competition between IF transitions and PS transitions because K_{OA} and K_{AH} also increase. To be more exact, K_{OA} increases 17.3 times and K_{OH} increases 10.6 times from $\epsilon_{\text{hinge}} = 1.0$ to $\epsilon_{\text{hinge}} = 0.0$. In contrast, in Φ_H model, the O-H transition process is slightly slowed down by the increase of the hinge flexibility, the corresponding fractional flux decreases also. The decrease of O-H transition and the slightly increase of the rate of O-A transition which is the rate-limiting step of PS routes together lead to the decrease of fractional flux of IF routes. Overall, the flexibility of hinge regions play an important role not only in the basin stability and but also in basin dynamics.

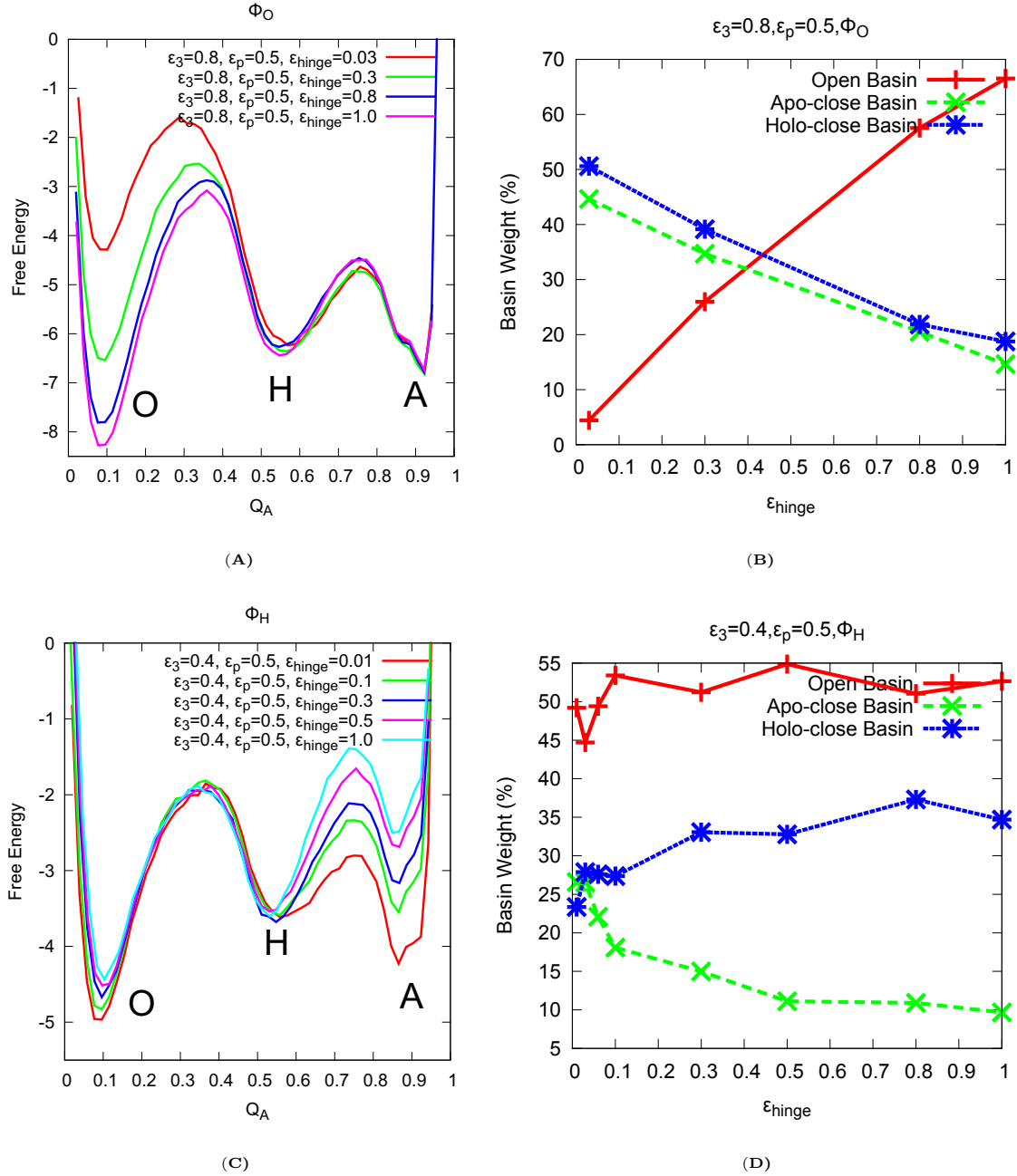


Fig. 8. Effects of ϵ_{hinge} on the thermodynamics with different native angle models. (A) and (C) show free energy profiles as a function of Q_A with different hinge flexibility with Φ_O model and Φ_H model. (B) and (D) show the correlations between basin probability, which reflects the stability for corresponding state, and hinge flexibility with Φ_O model and Φ_H model. For Φ_O model, decreasing of ϵ_{hinge} decrease the depth of open basin, for destabilizing the open-close basin and increasing the stability of apo and holo-close basin. But it has little influence on either $H \rightarrow A$ or $A \rightarrow H$ transition barrier. For Φ_H model, hinge flexibility mostly decreases the stability of holo-close basin and increases the stability of apo-close basin. However, there is little impact on $H \rightarrow O$ free energy barrier and stability of open basin. Overall, our results clearly show that the effect of hinge flexibility on protein stability. Although the models have different native biases, they all imply that increasing hinge flexibility can decrease the free energy barriers from their native basin to other basins.

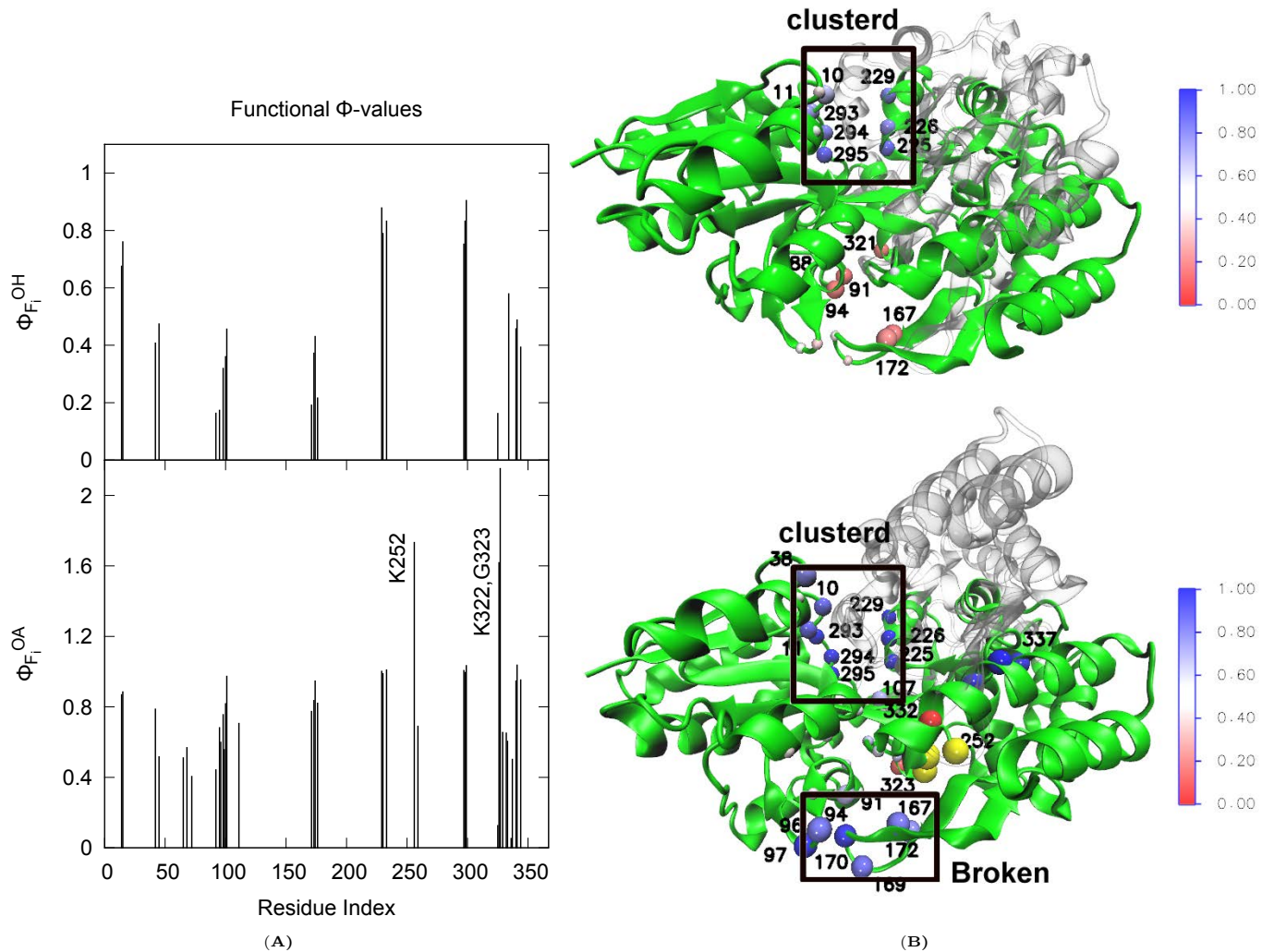


Fig. 9. Functional Φ_{F_i} values analysis of transition state ensembles TS_{OH} and TS_{OA} . (A) Functional $\Phi_{F_i}^{OH}$ and $\Phi_{F_i}^{OA}$ as a function of residue index. (B) Functional $\Phi_{F_i}^{OH}$ and $\Phi_{F_i}^{OA}$ mapped onto the two-ending crystal structures. Upper subfigure shows the open and holo-close structures. Lower subfigure shows the open and apo-close structures. Open structure is represented by green cartoon. The CTD of apo- or holo-close structure is demonstrated by grey transparent cartoon. Note that the structures of open and close states are aligned according NTD. The residues with Φ_{F_i} values are represented by spheres with different colors according to the right color bar in which blue means high value and red means low value. The best targets for mutations to perturb the transition rates are those residues for which Φ_{F_i} close to either 0 or 1. Here, the residues with modest Φ_{F_i} values (0.4 to 0.6) are represented by small white spheres. In addition, if $\Phi_{F_i}^{RP}$ is larger than 1, the residue will be colored in yellow. From the TS_{OH} analysis, it indicates that residues in NTD (D10, K11, K293, P294, L295) and these residues in CTD (P225, W226, S229) play a key role in stabilizing the holo-close state ($\langle Q_i \rangle^H$ is larger than $\langle Q_i \rangle^O$) and contribute to closure of MBP ($\Phi_{F_i}^{OH} > 0.6$). In holo-close state, these residues are clustered together to form a tight interaction network which can be seen in contact probability map in Fig. 11, and thus close the tip of active site groove. Our model also predicts that residues F88, D91, R94, N96, Y167, N169, Y172 and Q321 are especially close to open state in TS_{OH} ($\Phi_{F_i}^{OH} < 0.2$). In addition, the $\Phi_{F_i}^{OH}$ values for residues Y95, N96, G170, K171 (white small spheres) are about 0.5 and $\langle Q_i \rangle^O$ values for them are larger than $\langle Q_i \rangle^H$, indicate that the interactions between the balancing loop and NTD are partially broken in TS_{OH} , however the interactions between D91, R94 and Y167, Y172 are not broken. From the TS_{OA} analysis, it indicates that the clustered residues are the same as in TS_{OH} except that K38 also plays an important role to stabilize the apo-close state. Furthermore, in helix in C2 domain, there are not only high Φ_{F_i} residues (W336, Y337, R340), but also low Φ_{F_i} residue (M332). We found $\Phi_{F_i}^{OA}$ values of K252, K322 and G323 are larger than 1 (yellow spheres) caused by that $\langle Q_i \rangle^{OA}$ is out of the range of $\langle Q_i \rangle^O$ and $\langle Q_i \rangle^A$. Indeed, for K252, $\langle Q_i \rangle^O$, $\langle Q_i \rangle^{OA}$ and $\langle Q_i \rangle^A$ are 3.8, 2.8, 3.3 respectively. For K322, they are 3.9, 2.6, 3.1. And 2.7, 1.6, 2.2 for G323. K252 is located at the loop region of linker1, and K322 and G323 are located at the loop region of linker2 (balancing linker).

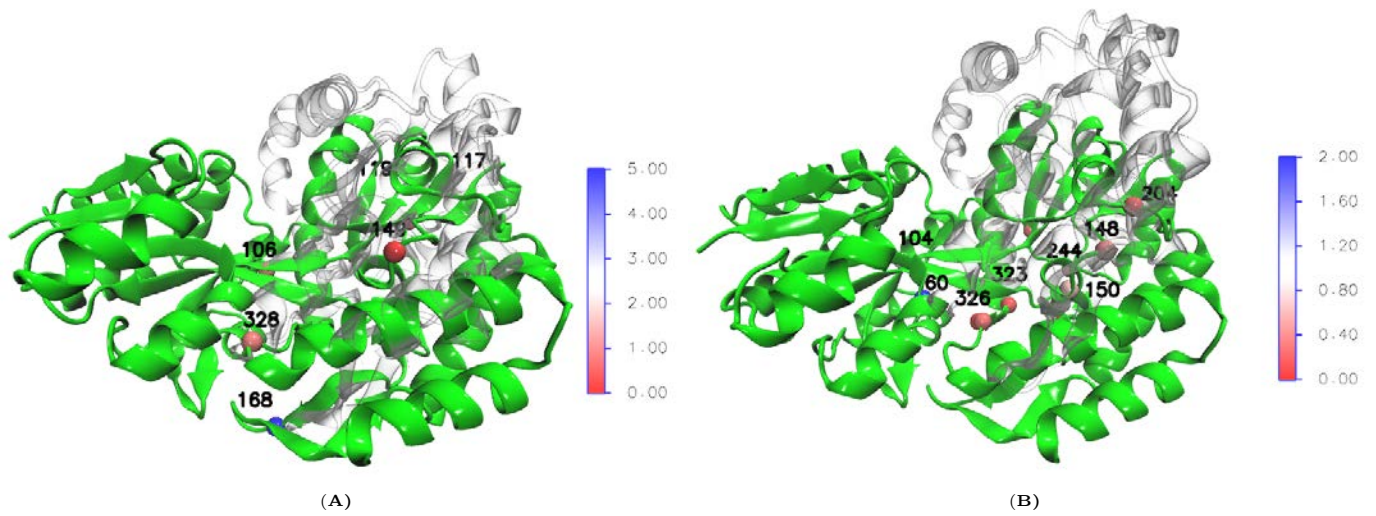
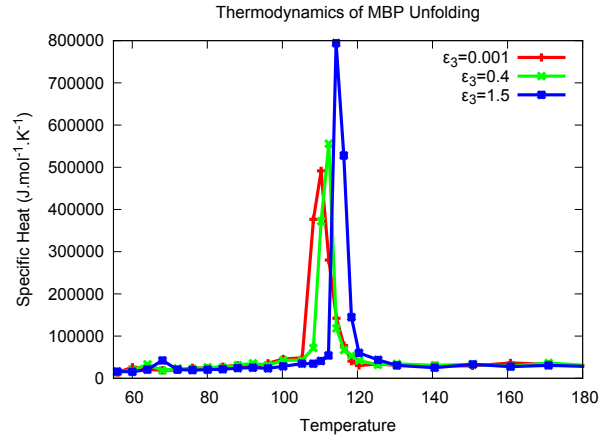


Fig. 10. Local Cracking for open-to-close transitions. (a) Unfolding points at transition state of O-H transition. (b) Unfolding points at transition state of O-A transition. Our model indicates that the local unfolding points may contain V106, L117, P119, E149, E168 and N328 in TS_{OH} . And for TS_{OA} , these unfolding region may include H60, I104, P150, Q148, T204, P244, K322, G323, M326. These residues are located at the loop regions with high flexibility with exception of I104 and V106 which are located in linker1 and approach to the known hinge residue E107. Especially, the high Ψ_i value for E168 which is located at balancing loop may be caused by the high flexibility of balancing loop due to not interacting with NTD. In addition, N328 in TS_{OH} and K322, G323, M326 in TS_{OA} are located at the loop region of balancing linker.



(A)

Fig. 11. Effects of ligand concentration on specific heat curves. The parameters are $\epsilon_1 = 1.0, \epsilon_2 = 0.4, \epsilon_o = 1.0, \epsilon_a = 1.3, \epsilon_h = 0.2, \epsilon_p = 0.5, \epsilon_{hinge} = 0.01$, with Φ_H . A shift of the peak of the specific heat curves to higher temperatures is observed when increasing ϵ_3 . This suggests the increase of protein stability. Along with the increase of stability, the folding cooperativity increases as well. This is evident from the narrowing of the specific heat profile when increasing ϵ_3 . From the results in simulations by our model it reveals that, under higher ligand concentration or with stronger ligand-binding interactions, the ligand binding links the two domains more tightly, leading to higher cooperativity, as well as stability.

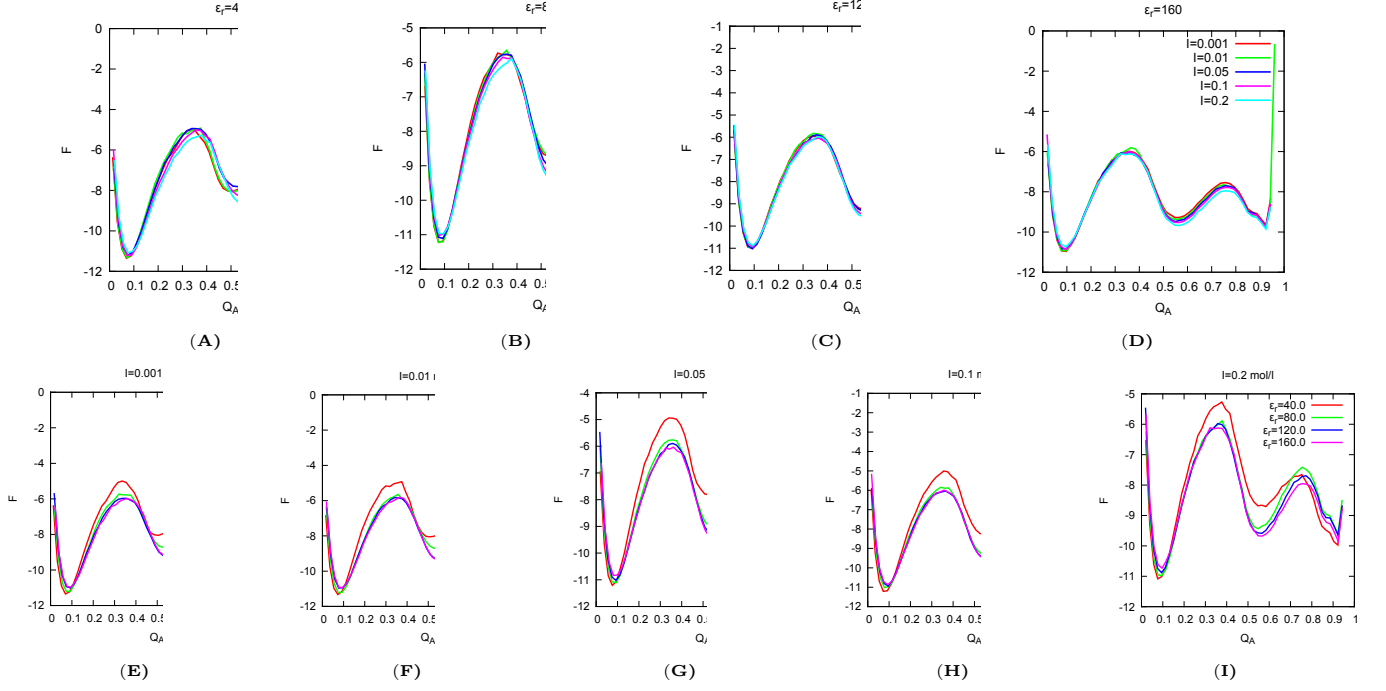


Fig. 12. Test of ϵ_r and ion strength. $\epsilon_1 = 1.0, \epsilon_2 = 0.4, \epsilon_o = 1.0, \epsilon_a = 1.4, \epsilon_h = 0.2, \epsilon_3 = 0.8, \epsilon_p = 0.5, \epsilon_{hinge} = 1.0$, with Φ_O .