



Ligand-bound glutamine binding protein assumes multiple metastable binding sites with different binding affinities

Lu Zhang ^{1,5}✉, Shaowen Wu^{2,5}, Yitao Feng^{2,5}, Dan Wang², Xilin Jia^{1,3}, Zhijun Liu⁴, Jianwei Liu² & Wenning Wang ²✉

Protein dynamics plays key roles in ligand binding. However, the microscopic description of conformational dynamics-coupled ligand binding remains a challenge. In this study, we integrate molecular dynamics simulations, Markov state model (MSM) analysis and experimental methods to characterize the conformational dynamics of ligand-bound glutamine binding protein (GlnBP). We show that ligand-bound GlnBP has high conformational flexibility and additional metastable binding sites, presenting a more complex energy landscape than the scenario in the absence of ligand. The diverse conformations of GlnBP demonstrate different binding affinities and entail complex transition kinetics, implicating a concerted ligand binding mechanism. Single molecule fluorescence resonance energy transfer measurements and mutagenesis experiments are performed to validate our MSM-derived structure ensemble as well as the binding mechanism. Collectively, our study provides deeper insights into the protein dynamics-coupled ligand binding, revealing an intricate regulatory network underlying the apparent binding affinity.

¹State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian, China.

²Department of Chemistry, Institutes of Biomedical Sciences, Multiscale Research Institute of Complex Systems, Fudan University, Shanghai, China.

³University of Chinese Academy of Sciences, Beijing, China. ⁴National Center for Protein Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. ⁵These authors contributed equally: Lu Zhang, Shaowen Wu, Yitao Feng. ✉email: luzhang@fjirsm.ac.cn; wnwang@fudan.edu.cn

Elucidation of protein dynamics and understanding how it is coupled with molecular recognition offer great insight into molecular control of cellular function^{1–7}. Although the modulation role of protein dynamics in ligand binding has seen much attention, a full microscopic description of the coupling between conformational dynamics and ligand binding is still challenging. Fundamentally, protein structural plasticity entails plenty of metastable states of protein in both ligand unbound and ligand-bound forms, which may not be easily accessible by experimental techniques. The diversity of structural features, energetics and kinetics of these states could result in binding thermodynamics and kinetics far more complicated than the simple models such as induced-fit and conformational selection^{8–16}. Therefore, dissecting the detailed conformational dynamics and energetics underlying ligand recognition of proteins is critical for understanding the modulation of affinity and related drug design.

Periplasmic binding proteins (PBPs) provide a typical model system to investigate the impact of protein dynamics on the ligand binding¹⁷. They are a superfamily of proteins responsible for binding and delivering substrates across the cell membrane by collaborating with the cognate ATP-binding cassette transporters^{17,18}. X-ray crystallographic studies, nuclear magnetic resonance (NMR), single molecule fluorescence resonance energy transfer (smFRET) and phosphorescence spectroscopy have been applied to investigate the conformational dynamics of PBPs^{13,19–31}. Glutamine binding protein (GlnBP) is a PBP that assists L-glutamine uptake in *Escherichia coli* and it has a similar three-dimensional structural scaffold to other PBPs^{18–20,32–37}. Crystal structures^{19,20} and NMR paramagnetic relaxation enhancement studies^{25,28,30} have suggested that ligand-bound GlnBP only adopts closed conformation while ligand-free GlnBP is solely in the open conformation, supporting the conventional Venus Fly-trap model. However, phosphorescence spectroscopy, NMR residual dipolar couplings, and smFRET experiments suggested inter-domain dynamics in ligand-free GlnBP^{24,30,31}, raising the possibility that conformational selection could play a role in its molecular recognition. Complementary to experimental techniques, molecular dynamics (MD) simulations and Markov state models (MSMs) method have become popular tools to decipher the dynamic structural ensemble of proteins at both atomic resolution and biologically relevant timescales^{38–53}. They have been adopted to discover metastable state and provide the kinetic information for molecular recognition for PBPs^{31,54–58} and elucidated multiple metastable states for ligand-free GlnBP³¹. Moreover, computational coarse-grained model proposed GlnBP could undergo large-scale closed-to-open transition⁵⁹. These studies demonstrate the conformational complexity of GlnBP and pave the way for elucidating its ligand binding mechanism. However, there are still critical questions remain largely unsolved and exploring the conformational dynamics of ligand-bound-GlnBP is prerequisite for fully understanding its molecular recognition mechanism.

In this work, we have combined MSM based on extensive MD simulations, smFRET, and site-directed mutagenesis to obtain a comprehensive insight into the conformational dynamics of ligand-bound GlnBP and elucidate its role in molecular recognition. Both our experimental data and computational models suggest that ligand-bound GlnBP has complex conformational dynamics in solution. Besides the inter-domain motion implicated by the crystal structures, we have also observed ligand migration between two domains and found a binding pocket in the small domain. The inter-state transitions estimated by the kinetic network model, as well as the different binding affinities of multiple conformational states together suggest an intricate interplay between the conformational dynamics and ligand binding, which were further validated by site-directed

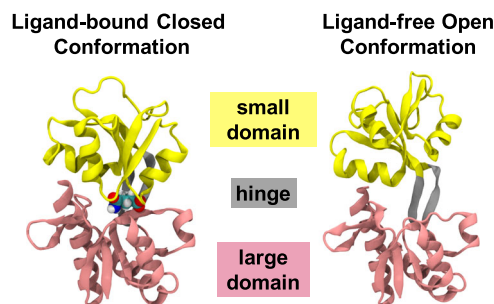


Fig. 1 Crystal structure of the ligand-bound closed and ligand-free open GlnBP. The small domain, the large domain and the hinge is shown in yellow, pink and grey, respectively. In the ligand-bound closed conformation, the ligand is shown in spheres.

mutagenesis and smFRET experiments. Taken together, our work offers not only a deeper insight into the complex conformational diversity of ligand-bound GlnBP, but also provides direct evidence that protein dynamics is essential for ligand binding.

Results

Ligand-bound GlnBP shows multiple ligand binding sites.

Crystal structures have demonstrated that GlnBP is composed of two globular domains, the large domain (protein residues 5–84 and 186–224) and the small domain (protein residues 90–180), connected by a hinge region (protein residues 85–89 and 181 and 185), and the substrate glutamine binds at the domain surface (Fig. 1)^{19,20}. However, our MD simulations of ligand-bound GlnBP have sampled a broad conformational space beyond the crystal structures. The simulation was initiated from the crystal structure²⁰ of ligand-bound GlnBP (closed conformation), followed by several rounds of simulations (~60 μ s) to reach a converged conformational space (Supplementary Fig. 1a, see “Methods” for details). To gain insight into the underlying thermodynamics and kinetics, we constructed the MSM^{38–46} based on the MD trajectories using a recently developed algorithm time-structure-based independent component analysis (tICA)^{44,45} (see “Methods” for details). In particular, the MD conformational space was decomposed into ~700 microstates according to the kinetic similarity. To examine the underlying molecular mechanism, the microstates were further grouped into eight macrostates, which represent local minima in the free energy landscape and their inter-state transitions are separated by energy barriers. Besides the two states similar to the crystal structures^{19,20}, another six macrostates were elucidated, which are well separated by not only the open-closed transitions, but also the ligand displacement between two domains (Fig. 2). The extent of opening was measured by the Ca–Ca distance between Thr59 (large domain) and Thr130 (small domain) (Fig. 2b). These two residues were chosen as they were used to attach fluorophore labels in the smFRET measurements (see below for details). Four of the eight states (S1, S6, S1', and S4') exhibit short average distances between 35 and 36 Å, and could be assigned as closed conformations. The S2 and S5 states could be assigned to open conformations due to the long average distance (43.1 Å for S2 and 46.1 Å for S5). The inter-residue distances of states S3 and S4 are moderate and thus were assigned as the “semi-closed” states. Examination of the representative structures of the eight macrostates revealed that the ligand could bind at different sites on GlnBP. In state S4, S5, and S6, the binding site is at the large domain. In states S1, S2, and S3, the ligand binds at the small domain (Fig. 2a), while the ligand binds at the interface between the large and small domain in states S1' and S4'. This assignment

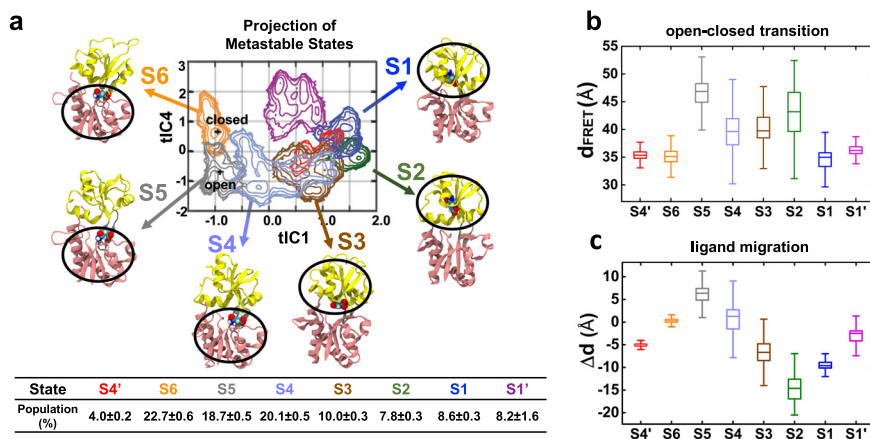


Fig. 2 MSM analysis elucidates eight macrostates, demonstrating not only the open-closed transitions, but also the ligand displacement between two domains. **a** Projection of macrostates onto two tICs. Representative conformations of the six states on the major transition pathway are shown with the ligand-binding region highlighted in black circle. MSM-derived populations of macrostates are labeled. The crosses label the projection of the crystal structures (Fig. 1). **b** The open-closed transition of protein was measured as distance between C α atoms of Thr59 and Thr130. **c** The ligand migration was estimated by its relative distance to the large/small domain. The averaged pairwise distances between the heavy atoms of the ligand and the C α atoms in each domain were calculated. Δd equals to the $d(\text{ligand-small domain}) - d(\text{ligand-large domain})$. In **b**, **c**, the box plot for each macrostate was calculated based on all the MD conformations belonging to the specific state.

of binding site was further confirmed by the relative distance between the ligand and two domains (Fig. 2c). Interestingly, we notice that although the tICs generated by tICA^{44,45} are usually a mixture of various conformational changes, the ligand migration and protein opening-closing motion were well captured by the first and the fourth tICs (Fig. 2a), suggesting that these conformational changes indeed make obvious contributions to the ligand-bound GlnBP dynamics.

Ligand-binding affinity correlates with protein dynamics. The ligand binding affinities for different macrostates were estimated by Poisson–Boltzmann surface area (MM-PBSA) method^{60,61} (see “Methods” for details). This method has been widely applied in areas of biomolecular studies and computational drug design to estimate the relative binding free energies^{62,63}. We note the binding free energy of individual macrostate could not be measured by experimental method, but it can shed light on the relative binding strength between protein and ligand in each state. Accordingly, we found the ligand at large domain has higher binding affinity than it is at small domain (Fig. 3a). This matches with the observation that large-domain binding states S4, S5, and S6 have higher populations than the small-domain binding states S1, S2, and S3 (Fig. 2a). The calculated ligand binding affinities are also in line with the analysis of the ligand–protein interactions (Fig. 3). Ligand in S5 and S6 states show similar interactions with the large domain: the ligand’s alpha group forms stable interactions with the large-domain residues Gly68, Thr70, and Arg75 (averaged distance ~ 3.0 Å); the ligand’s side chain form a bit stronger interactions with Asp10 and Ala67 in S6 state than in S5 state (Fig. 3b). However, S6 has remarkably higher binding affinity than S5 (Fig. 3a), due to the binding site of S6 state is on the large-domain surface and additional interactions of the ligand with the small-domain residues Asp157 and Gly119 can be formed when protein adopts closed conformation (Fig. 3c). On a sharp contrary, S1 and S2 demonstrate different protein conformations (Fig. 2b), but they have similar ligand binding affinity (Fig. 3a). This can be explained by the observation that their binding pockets in the small domain are relatively buried (Supplementary Fig. 2) and ligand only interacts with protein residues in the small domain in these two states. Specifically, ligand in S1 and S2 states shows similar interactions with small-domain

residues Ala113, Lys115, Gly119, Leu155, and Ser120 (Fig. 3c), while does not interact with the large-domain residues (Fig. 3b). Thus, the opening-closing motion would not perturb the ligand–protein interactions in S1 and S2 states. These observations altogether explain why S1 and S2 have similar ligand binding affinity (Fig. 3a), even though they demonstrate different protein conformations (Fig. 2b).

Furthermore, we found different binding sites can give rise to diverse binding affinities. As shown in Fig. 2b, S2 and S5 states show d_{FRET} of 43.1 ± 3.9 Å and 46.1 ± 3.2 Å, respectively, suggesting that their opening extent are similar. However, the binding affinities of S2 and S5 are notably different, with the binding affinity of S5 state obviously higher than that of S2 state (Fig. 3a). How these states are kinetically connected is a question worthy of exploring and this investigation would assist the understanding of the interplay between the protein dynamics and ligand binding.

Protein dynamics modulates the kinetic transition network.

The kinetic transition network was elucidated by MSM (Fig. 4a), according to which the macrostates were roughly grouped into two clusters: one involves S1, S2, and S3 with ligand at the small domain, and the other involves S4, S5, and S6 with ligand at the large domain. The two clusters are connected by S3 and S4. S1’ and S4’ are two off-pathway states that only connect to S1 and S4, respectively. We found that the opening-closing transitions within cluster (S1 \leftrightarrow S2 or S5 \leftrightarrow S6) occur very fast (<10 μ s), and there is no intermediate state during the transitions. Therefore, GlnBP could undergo quick opening-closing movement with the ligand bound at either the large or small domain. On the other hand, transitions through the semi-closed states S3 and S4 are slower, especially the S3 \rightarrow S1 and S3 \rightarrow S2 transitions. This suggests that besides the opening-closing motions, other conformational changes also determine the kinetics.

First, we found that the kinetics is partially determined by the structural rearrangements involving helix V (residues 138–146), helix VI (residues 160–168) and strand J (residues 111–115) in the small domain (Fig. 4b). The extent could be measured by the value of $\Delta d = d_1 - d_2$ (d_1 for the space between helix VI and helix V; d_2 for that between helix V and strand J). As shown in Fig. 4b, in the semi-closed S3 and S4 states, the values of Δd are moderate.

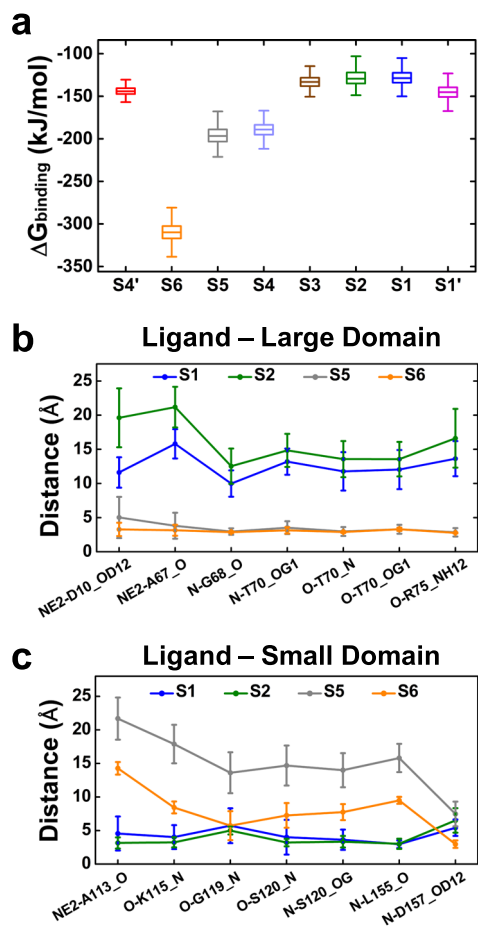


Fig. 3 Ligand-binding affinity is correlated with the ligand-protein interactions. **a** Ligand-binding free energies of macrostates are estimated by MM-PBSA method. We performed 1000 interactions of calculations to achieve the statistics about the reported data. In each interaction, one random conformation was selected from each microstate and the binding free energy of each macrostate was estimated by the population weighted sum of the binding free energies of all the microstates belonging to the specific macrostate. Box plot represent the data distribution resulted from the 1000 times of calculations. **b** Interactions between ligand and the protein residues in the large domain. **c** Interactions between ligand and the small-domain residues. In **b-c**, the atoms used for distance calculations are labeled next to the x-axis and the minimum distance is reported when there are chemically equivalent atoms. The means and standard deviations of distances for each macrostate were calculated using all the MD conformations belonging to the specific state.

However, in S5, S6, and S4' states, Δd values are small, implying the binding pocket of the small domain is almost locked. On the contrary, S1, S2, and S1' states have obviously larger Δd values, thereby that pocket is wide enough to fully accommodate the ligand. This structural rearrangement plays an important role in not only modulating the kinetics but also to make enough space for ligand to diffuse into the binding site on the small domain.

Further structural analysis reveals the flexibility of the hinge region also tunes the kinetics. The rigidity of the hinge region is mainly contributed by four pairs of hydrogen bonds (Fig. 4c). As ligand migrates from large domain to small domain, the hinge becomes more flexible, mainly manifested by the decreasing bonding strength between the backbone of Y185 and that of Y85/Y86. We also notice although the off-pathway state S1' has similar hydrogen bond number with S1 state, their hydrogen bonding patterns are different, involving weakening the hydrogen bond

(S88_O-Q183_N) and simultaneously strengthening the hydrogen bond stability between Y86_O and Y185_N. This also rationalizes the slow off-pathway transition between S1 and S1' states. The other off-pathway transition S4→S4' is also relatively slow and this is due to the variation of Δd (Fig. 4b) as well as the expansion of a loop between strand J and helix IV to allocate the ligand (Supplementary Fig. 3).

Overall, by combining the kinetics derived from MSM analysis as well as detailed structural analysis, we have pinpointed the key structural features that determine the transition kinetics, providing the basis to understand the molecular recognition and manipulate the binding affinity of ligand-bound GlnBP.

A concerted model for molecular recognition. To understand the ligand binding mechanism, a comparison between ligand-free and ligand-bound GlnBP was made. The conformational space of ligand-bound GlnBP explored herein shows that ligand binding expands the conformational space with respect to that of ligand-free GlnBP³¹ (Supplementary Fig. 4). First, we performed principal component analysis based on all C α atoms using all the MD conformations of ligand-bound GlnBP and the same eigenvectors were adopted for projecting the ligand-free GlnBP conformations³¹. The projection of each macrostate sampled in ligand-bound GlnBP was compared to the overall projection of ligand-free GlnBP. Projection of S5 and S6 onto the top two components of principal component analysis shows considerable overlap with that of ligand-free GlnBP; while other states are partially (S2, S3, and S4) or not sampled (S1, S1', and S4') by ligand-free GlnBP (Supplementary Fig. 4a). Second, as the ligand binding induces the conformational rearrangement in the small domain (Fig. 4b), we also projected macrostates of ligand-bound GlnBP as well as ligand-free MD conformations onto the dimension of Δd . The comparison shows that states S3, S4, S5, and S6 remarkably overlap with ligand-free GlnBP, while the other states (S1, S1', S2, and S4') populate distinct regions (Supplementary Fig. 4b). Taken these results together, we may conclude that four macrostates (S3, S4, S5, and S6) of ligand-bound GlnBP have their counterparts in ligand-free GlnBP. The other four macrostates (S1, S1', S2, and S4') identified in ligand-bound GlnBP, which involve structural rearrangements of the small domain and/or remarkable inter-domain twist, are absent or very rarely populated in ligand-free GlnBP.

Based on the structural features, the kinetic network and the comparison between ligand-bound and ligand-free GlnBP discussed above, we proposed a concerted mechanism of ligand binding (Fig. 5). Although the S6 state has the highest binding affinity and its conformation exists in the ligand-free GlnBP, it is inaccessible for initial ligand binding, since the two domains are locked up and has limited space to allow ligand to diffuse in. This is also reflected by the smallest solvent accessible surface area (SASA) of S6 (Supplementary Fig. 2). On the contrary, the open state S5 together with the two semi-closed states S3 and S4 are sampled by ligand-free GlnBP and have remarkably larger SASAs (Supplementary Fig. 2), serving as the potential conformations for initial ligand binding. The subsequent transitions have several possibilities following the kinetic network (Fig. 4a). In this scenario, “conformational selection” works for the initial selective binding to S5, S4 or S3, and “induced fit” corresponds to subsequent transitions to the closed state S6 or states that barely exist in the absence of ligand (Fig. 5). In this regard, our studies provide another case that two mechanisms are not mutually exclusive but can be collaborative for ligand binding^{57,64,65}.

Experimental examination of the conformational dynamics. One major finding in this study is ligand-bound GlnBP has

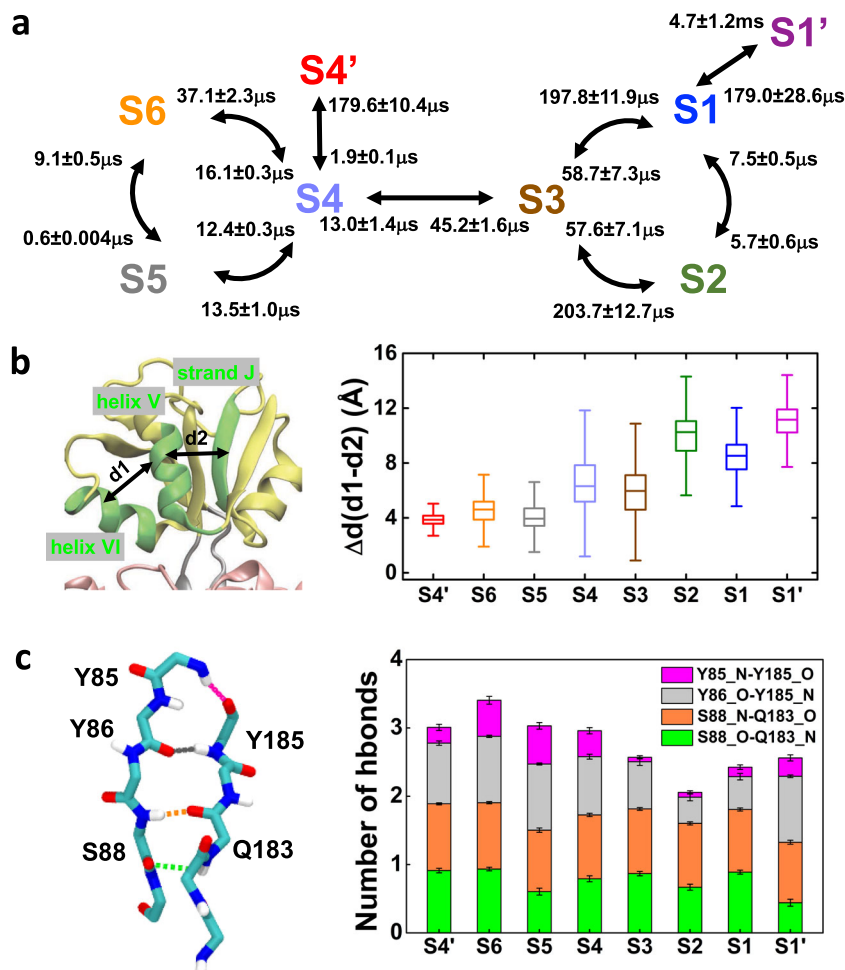


Fig. 4 Modulation of the kinetic transition network by essential protein conformational changes. **a** Transition pathways elucidated by MSM, with mean first passage times alongside the arrows. **b** Structural rearrangement in the small domain manifested in Δd . d_1 denotes the distance between helix VI (residues 160–168) and helix V (residues 138–146), while d_2 for the helix V-strand J (residues 111–115) distance. Box plot for each macrostate represents the distribution of Δd . All the MD conformations belonging to the specific state were included in the statistics. **c** The rigidity of the hinge (residues 85–89 and 181 to 185) is represented by four pairs of hydrogen bonds. Each pair is shown with an individual color in the stacked bar. Bootstrapping algorithm with replacement was applied to estimate the statistical errors of the number of hydrogen bonds, in which 100 samples were generated and each sample contains 100 random conformations from the macrostate under investigation.

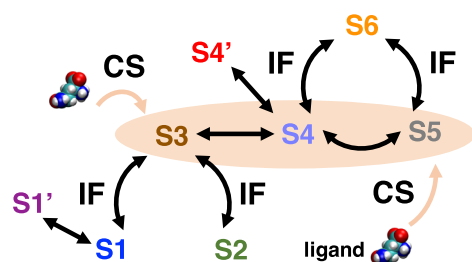


Fig. 5 Ligand-binding mechanism for GlnBP. “Conformational selection” (CS) works for the initial selective binding to S_5 , S_4 or S_3 state, followed by “induced fit” (IF) for subsequent transitions to the states that barely exist in the absence of ligand.

multiple conformational states other than the fully closed conformation captured by the crystal structure. To examine if there exist multiple conformational states for ligand-bound GlnBP in solution, we performed smFRET experiments to measure its protein conformational dynamics (see “Methods” for details). Thr59 and Thr130 (the same residues used above for measuring the inter-domain distance in Fig. 2b) were mutated to

cysteines and labeled with fluorophores (Alexa fluor 555-maleimide or Alexa fluor 647-maleimide) (Fig. 6a). We found FRET efficiency displays a broad distribution (Fig. 6b, c), suggesting the ligand-bound GlnBP can adopt other conformations other than the closed one as suggested by the crystal structure. Four states could be identified from transition density plot by hidden Markov model (HMM) state analysis (Supplementary Figs. 5–6, see “Methods” for more details). They are named as E1, E2, E3, and E4, with the efficiency values centered at 0.15, 0.31, 0.51, 0.68, respectively.

We note a quantitative comparison between the MD simulation and smFRET measurement is not feasible due to both computational limitations (e.g., not including the fluorophores in the simulation system) and many aspects that affect the experimental FRET efficiencies^{66,67}. Moreover, the four states (E1 to E4) along the inter-residue distance or FRET efficiency do not correspond to the macrostates of MSM, since this single degree of freedom could not discriminate all the eight macrostates. In this regard, we only made a rough comparison using overall distribution of the inter-C α distances between the dye-labeled residues. The calculated distance distribution can be well fitted with four Gaussian functions (Supplementary Fig. 7),

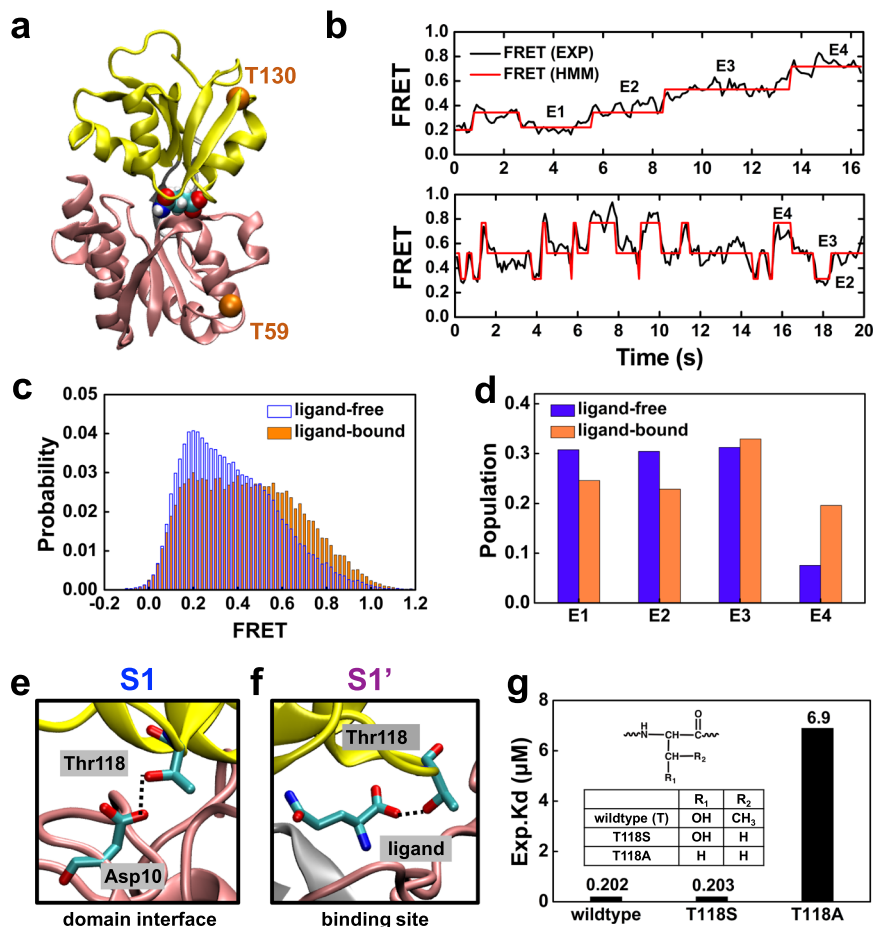


Fig. 6 Experimental investigation of protein dynamics by smFRET and mutagenesis. **a** GlnBP structure with the orange spheres denoting the positions where fluorophores labels are attached. **b** smFRET time traces for the ligand-bound wildtype GlnBP. **c** smFRET distribution for ligand-bound (orange) and ligand-free (blue) GlnBP. **d** Populations of the four states identified in smFRET for both ligand-free and ligand-bound states. **e** In S1 state, Thr118's side chain is interacting with Asp10 on the opposite domain. **f** In S1' state, Thr118's side chain forms direct interaction with the ligand. **g** Ligand-binding affinity for wildtype, T118S and T118A. The chemical structures of wildtype and mutants are shown in the inset.

centered at 4.7 Å (21%), 3.9 Å (46%), 3.6 Å (22%), and 3.4 Å (11%), the populations of which are in general agreement with the populations of the four ligand-bound states (24%, 23%, 33% and 20% for E1, E2, E3 and E4, respectively (Fig. 6d)) in smFRET measurements.

Furthermore, comparison of FRET efficiency distribution between ligand-bound and ligand-free GlnBP shows a clear population shift upon ligand binding (Fig. 6c). The state population of E4 is obviously raised while simultaneously those of E1 and E2 are reduced (Fig. 6d and Supplementary Fig. 6). One possible explanation for this population shift is that the closed S6 state shows higher ligand binding affinity as well as stronger protein–ligand interactions than the open S5 state (Fig. 3a, b), therefore promoting the corresponding population shift to the more closed conformation. An alternative interpretation is that the ligand binding has induced the appearance of four states (S1, S1', S2, and S4') that are not existing or rarely sampled in the absence of ligand as discussed above. Three of them adopt closed conformations, with the total populations (~24%) almost six times than that of the open one (~4%) (Fig. 2a, b). The abundance of additional closed conformations induced by ligand-binding also contributes to the population shift.

Overall, consistent with our computational model, smFRET measurements also indicate that ligand-bound GlnBP has high conformational heterogeneity in solution and ligand binding has

induced the protein population shift to the more closed conformations.

Another major finding in this study is that ligand can bind in a buried pocket in the small domain. The most straightforward way to validate the small domain-binding site is to make mutations on the residues directly interacting with the ligand in these states. However, we notice these interactions are mainly through the backbone of the protein, instead of the side chain (Fig. 3c). Therefore, mutations would hardly make obvious perturbation to the binding affinity. Further structural investigations reveal that Thr118 may come into play to affect the protein's conformational dynamics as well as the ligand binding affinity. In the S1 state, the hydroxyl group in the Thr118's side chain forms stable hydrogen bonds (averaged distance ~3.1 Å) with the Asp10 in the opposite domain to assist locking the protein in the closed conformation (Fig. 6e). In the off-pathway state S1', Thr118's side chain directly interacts with the ligand's alpha group, helping to stabilize ligand inside protein (Fig. 6f). MM-PBSA calculations estimate that Thr118 contributes ~10% to the binding free energy of S1' state (see “Methods” for details). Based on these observations, we anticipate that Thr118 may affect conformational dynamics as well as ligand binding affinity.

Accordingly, we performed site-directed mutagenesis on Thr118 and measured the dissociation constants for the mutants (Fig. 6g, see “Methods” for details). Specifically, we designed

T118S mutation to maintain the hydroxyl group in its side chain, and T118A mutation to eliminate the hydroxyl group. The results demonstrate T118S mutation has a similar binding affinity ($K_d = 203$ nM) to the wildtype ($K_d = 202$ nM). On the contrary, T118A mutation greatly decreases (~ 35 folds) the ligand binding affinity ($K_d = 6.9$ μ M). These mutants underline the hydroxyl group of Thr118 is critical for ligand binding. It's also worthy to note that Thr118's potential function in modulating ligand binding affinity is pinpointed by the two extra conformational states from MSM. Therefore, the above experimental validation of Thr118's effect could also serve as a proof of the multiple conformations elucidated in this work. Furthermore, it also provides direct evidence that perturbation of the conformational space can affect the binding affinity.

Discussion

Elucidation of the conformational dynamics is essential to understanding its interplay with the biological function, especially the mechanism of molecular recognition^{1–7}. “Conformational selection” and “induced fit” are two extreme models to explain the ligand binding mechanisms^{8–16}. Although two mechanisms describe different dynamic processes, they both emphasize the role of protein dynamics in ligand binding. Previous X-ray crystallographic studies found GlnBP in a closed conformation upon ligand binding, and in an open conformation without ligand. The absence of ligand-free closed crystal structure leads to the assertion that GlnBP follows “induced fit” mechanism upon ligand binding, as “conformational selection” mechanism assumes protein can adopt various conformations even without the existence of ligand. However, phosphorescence spectroscopy²⁴ and coarse-grained simulation⁵⁹ have captured the inter-domain dynamics for ligand-free GlnBP. Consistently, the closed conformation was identified as one major conformation for ligand-free GlnBP using integrated computational and experimental methods³¹. These raise doubts about the commonly assumed “induced fit” mechanism for GlnBP.

In this work, we have combined MD simulation, MSM and experimental methods to examine the role of conformational dynamics in ligand binding. We have found ligand-bound GlnBP exhibits more complicated conformational dynamics than that of ligand-free GlnBP, and ligand can bind sites other than that observed in crystal structure. According to the microscopic models of conformational selection and induced fit mechanisms, the case of GlnBP can be viewed as a complex combination of the two paradigms. We also performed experiments to investigate the protein dynamics of GlnBP. Consistent with the MD and MSM results, our smFRET measurement also suggests there exist multiple conformational states for ligand-bound GlnBP. Furthermore, we identified one residue T118 that plays important role in modulating the ligand binding affinity based on our computational model and our prediction was validated by site-directed mutagenesis and binding affinity measurement. MSM constructed based on extensive MD simulations of GlnBP with the specific mutation would help to elucidated the detailed variation of protein dynamics caused by the mutation. It is noted there exists a large discrepancy between the ensemble-averaged absolute binding free energy ($\Delta G_{\text{binding}}$) calculated using MM-PBSA method (-197.1 ± 4.1 kJ/mol) and that obtained from the ITC measurement (202 nM, corresponding to -63.0 kJ/mol at $T = 291.15$ K). One major source of this discrepancy is the conformational entropy, which is not considered in the MM-PBSA calculations^{60,62}. Since the entropy's contribution in different metastable states with the same ligand are usually similar⁶², it is reasonable to leave out the entropy term in the calculations when the relative binding affinity rather than the absolute value is

needed. Actually, MM-PBSA is popular in estimating the relative binding free energies ($\Delta(\Delta G_{\text{binding}})$) in many applications with reasonable accuracy and efficiency^{62,63,68}, which fits our purpose to correlate the relative ligand binding affinity of each conformational state to the protein dynamics in the current study. Altogether, these computational and experimental results suggest the complex conformational dynamics of ligand-bound GlnBP. More importantly, they provide direct evidence that binding affinity can be manipulated by regulating protein dynamics.

Complex conformational dynamics and multiple binding sites have also been found in other protein–ligand recognition systems, in sharp contrast to the simple Venus Fly-trap model. This feature might be a common paradigm of molecular recognition adopted by many proteins that have yet to be discovered. The obvious functional advantage is that the complex dynamics entails a complex kinetic network of ligand binding, which allows more flexible regulation of binding affinity. It is worth noting that multiple conformational states and their transition kinetics are very difficult to characterize experimentally. The strategy of combining experimental and computational methods shows its strength in elucidating the microscopic mechanism of molecular recognition.

Methods

MD simulations. Set up of MD simulation model: The ligand-bound crystal structure (PDBID: 1WDN¹⁹) was used as the structural basis for model construction. The terminal residues Lys4, Pro225, and Lys226 were removed to make the residue number consistent with the open crystal structure (PDBID: 1GGG²⁰). The system was solvated in a dodecahedron water box with sodium chloride concentration of 0.3 mol/L. To consider the protein may transit to the open conformation, the box size was determined by the open crystal structure, making box edges at least 10 Å from the protein surface. Protein was solvated with TIP3P water molecules⁶⁹ and counter-ions were also added to neutralize the system. AMBER03 force field parameters⁷⁰ are used for simulating protein and ions. Partial charges of ligand were derived following RESP scheme^{71,72} and other parameters were extracted by referring to those of glutamine in the AMBER03 force field. First, 10,000-steps energy minimization was performed. Next, the system was simulated for 200 ps with position restraint on all the heavy atoms with a force constant of $10 \text{ kJ} \times \text{mol}^{-1} \times \text{Å}^{-2}$ under NVT ensemble ($T = 298$ K), followed by another 500 ps position restraint under NPT ensemble ($T = 298$ K, $P = 1$ bar). Afterwards, we removed the restraint and performed a 10 ns NPT simulations, the final configuration of which was used to initiate 20 independent 100 ns production NVT ($T = 298$ K) simulations with different initial velocities. In the simulations, we applied V-rescale thermostat⁷³ with the coupling time constant of 0.1 ps. A cutoff of 12 Å was used for Lennard-Jones interactions. The long-range electrostatic interactions beyond the cut-off at 12 Å were treated with the Particle-Mesh Ewald (PME) method⁷⁴. The neighbors list was updated every 10 steps. An integration time step of 2.0 ps was used and the LINCS algorithm⁷⁵ was applied to constrain all the bonds. We saved the snapshots every 50 ps. All the simulations were performed using Gromacs 2016.4^{76,77}.

Seeding unbiased MD simulations: Several rounds of simulations were performed by seeding from different regions of the conformational space.

We collected the MD conformations from the first round of 20×100 ns simulations after removing the first 10 ns data from each trajectory. The conformations were divided into 20 clusters by k-center clustering algorithm^{78–80} based on the r.m.s.d. of the 220 Ca atoms after aligning the conformations by the large domain (Ca atoms of residue IDs 5 to 84 and 186 to 224). One random conformation was extracted from each cluster and used as the seed for one 100 ns MD simulation in the second round. In the same way, we collected all the conformations from the second round of 20×100 ns simulations and clustered them into 20 clusters. One random conformation from each cluster was selected to initiate one 100ns trajectory as the third round of MD simulations.

To examine if the ligand can reside at the domain surface when protein is open, we constructed another two models. In the first model, we aligned the closed crystal structure (PDBID: 1WDN²⁰) to the open crystal structure (PDBID: 1GGG²⁰) by the large-domain residues. Ligand was then extracted from the aligned closed conformation and inserted into the ligand-free-open conformation (named as “LDbind-open”). In the second model, the alignment was made according to the small-domain residues (Ca atoms of residue IDs 90 to 180) and in this way the ligand was located at the small-domain surface (named as “SDbind-open”). Similar system set up as elaborated in the subsection “Set up of MD simulation model” was used herein and 10 independent 100 ns production NVT ($T = 298$ K) simulations were performed for each model. For the “LDbind-open” model, ligand keeps binding at the large-domain surface in nine simulations. However, for the “SDbind-open” model, none of the simulations shows ligand bound at the

small-domain surface and we observed ligand diffused from the small domain to the large-domain surface in two simulations. Therefore, we speculated the large-domain surface can serve as a ligand binding site when protein is in the open conformation. In this regard, we collected the conformations from the simulations with ligand binding at the protein (nine from the “LDbind-open” model and two from the “SDbind-open” model) after removing the first 10 ns from each trajectory, and did a 20-state K-center clustering and used the center conformation for seeding the second of simulations.

We combined the second and the third round of simulations from the ligand-bound closed crystal structure, as well as the second round of simulations from the two manually constructed models to do a K-center clustering and divided the conformations into 50 clusters. One random conformation was extracted from each cluster to seed the next round of simulations. In a similar way, we performed another eight rounds of MD simulations seeded from either the cluster center or one random conformation of each cluster, excepting the last two rounds in which two random conformations for each cluster were used for seeding. In total, we performed 610×100 ns MD simulations. In this work, we only focused on the case that ligand binds at the protein, therefore in each round of MD simulations, we examined the trajectories and removed those showing ligand dissociation before doing clustering. Finally, we have 576 MD trajectories for the MSM constructions.

Construction and validations of MSM. The advantage of MSMs is to facilitate the investigation of long-timescale dynamics based on many short MD simulations by discretizing both conformational space and time^{38–46,78,79,81–84}. Its basic idea is to partition the conformational space into metastable states, with each state corresponding to an energy minimum in the free energy landscape and the transition between them are slow. With a proper lag time Δt , fast motions within states are integrated out by discretizing time in units of lag time and the model becomes Markovian. That is, the probability for the system to visit a certain state at time $t + \Delta t$ is solely determined by its current position at t . Under the Markovian condition, the long-timescale dynamics can be obtained by propagating the transition probability matrix $T(\tau)$:

$$P(n\Delta t) = T(\Delta t)^n P(0), \quad (1)$$

where $P(n\Delta t)$ is the state population vector at time $n\Delta t$, and T_{ij} is the element of $T(\Delta t)$, denoting the transition probability from state i to state j after a lag time of Δt .

We followed a commonly used “splitting and lumping” procedure to construct the MSM.

Splitting conformations into microstates with validated MSM parameters: We adopted the tICA method^{44,45} combined with the K-center algorithm^{78–80} to divide the MD conformations. The dimensionality reduction by tICA has been found to capture the slow and biological relevant conformational changes^{44,45}. Optimal parameters for MSM construction were selected using a generalized matrix Rayleigh quotient (GMRQ)^{43,85} and parameters including atomsets, tIC number, tICA lag time and cluster number for K-center clustering were examined (Supplementary Fig. 8). Generally, GMRQ measures the ability of a model to capture the slowest dynamics and GMRQ scores, which are a summation over the several highest eigenvalues, serve as a good metric. The higher scores suggest the model is closer to the variational bound and is preferred. In practice, we performed 50 iterations of shuffle-split cross-validations. In each iteration, we randomly selected half number of trajectories as the training set for learning MSM and used the remaining trajectories as test set for scoring.

To consider both the inter-domain conformational change and the ligand diffusion inside the protein, we used the atomic pairwise distances between two domains, distances between the ligand and each domain, as well as the distance between the two FRET dyes inserted positions as the feature for tICA. To select the domain residues subject for the atom-pair calculations, we considered the residues within different cut off distances of the ligand. Both the ligand-bound-closed crystal structure and one representative closed conformation with ligand in the novel small-domain binding pocket were considered for the domain residue selection (Supplementary Fig. 8a–b). In the distance calculations, for the domain residues only Ca atoms were used, while for the ligand all the heavy atoms were included. In particular, we examined the cut off distances ranging from 5 to 11 Å and found GMRQ scores highest at 9 Å (Supplementary Fig. 8c). We also investigated tIC number, tICA lag times and cluster number for K-center clustering, and the parameters showing the highest GMRQ scores are selected (Supplementary Fig. 8d–f). Finally, we used 9 Å as the distance cutoff to select atom sets, 4 tICs, 10 ns as tICA lag time and divided the conformation into 750 microstates. To have better statistics within the microstates, we removed those microstates having fewer than 15 conformations (~1% of mean number of conformations per state) and got the 704 microstates for the subsequent model validation and analysis.

MSM validation: To determine the lag time Δt when the system becomes Markovian, we examine the implied timescales^{40,42,86} (τ_k), which is calculated as:

$$\tau_k = -\tau / \ln \mu_k(\tau), \quad (2)$$

where $\mu_k(\tau)$ is the k th eigenvalue of the transition probability matrix at lag time τ . If the model is Markovian, the implied timescale τ_k becomes constant regardless of the choice of τ ⁸⁷.

To take the uncertainties into account, we applied bootstrapping method by using 576 samples. For each sample, we randomly selected 576 trajectories with replacement from the ensemble of MD trajectories and calculated the implied timescales by constructing MSMs. Averaged values and errors were estimated from the 576 samples and reported in Supplementary Fig. 9a. In our model, we found the implied timescale plot reached plateau at around 30 ns, implying the system becomes Markovian after this time. Accordingly, we selected 30 ns as the lag time to construct our MSMs.

To further validate our model, we compared the residence probability (the probability for the system to remain in a certain microstate) as a function of time obtained from the propagation of the MSM with those from the original MD trajectories⁴¹ (Supplementary Fig. 9b) and they are in good agreement.

Examination of the convergence of conformational samplings and MSMs: In order to investigate the convergence of the MD conformational space, we compared the complete MD samplings (N rounds) with those excluding the last two rounds (N-2 rounds), and those excluding the last round (N-1 rounds) (Supplementary Fig. 1a). It's shown that no new metastable regions in the conformational space appears as the increase of the sampling, suggesting the free energy landscape has reached reasonable convergence. We also examined the robustness of our MSMs by constructing models from the three MD datasets. The results show the implied timescales are consistent with the increment of size of MD conformational ensemble (Supplementary Fig. 1b), reflecting that our conformational sampling is sufficient to predict the dynamics of the system.

Lumping and calculations of the quantities for metastable states: For visualization and revealing the molecular mechanisms, we further lumped microstates into eight metastable states by an improved Perron-cluster cluster analysis^{88,89}. To calculate the quantities for metastable states, we used the transition probability matrix constructed at microstate level to generate 100×300 ms Monte Carlo trajectories. The first 100 ms was used as equilibration and removed for the calculation of quantities. The stationary populations of metastable states (Fig. 2a) and the mean first passage time (Fig. 4a) for each transition were calculated by averaging the values over the 100 trajectories.

Estimation of binding free energies. We applied MM-PBSA method^{60,61} to calculate the binding free energies for each metastable state (Fig. 3a). We randomly selected 10 conformations from each microstate for the MM-PBSA calculations with SASA model for computing non-polar solvation energy. Default parameters are used for the calculations and the temperature is set to 298 K (MD simulation temperature). We performed 1000 trials to estimate the average and standard deviations for the binding free energy for each metastable state. In each trial, we randomly selected one conformation from each microstate and calculated its binding free energy; the binding free energy of one metastable state is then calculated as the weighted sum over those of all the microstates belonging to the corresponding metastable state, and the weight is determined by the relative population of the microstate in the respective metastable state.

The MM-PBSA method was capable to estimate the individual residue's contribution to the overall binding energy. The averaged binding energy from T118A was derived in a similar way as illustrated above.

The ensemble-averaged binding free energy contributed by all metastable states was computed as the weighted sum of the binding free energy of each microstate. The weight is determined by the population of the respective microstate. Mean and standard deviation were estimated by performing 1000 iterations of such calculations.

Protein preparation. *Escherichia coli* glnH was cloned from genomic DNA of *Escherichia coli* BL21(DE3) strain by standard PCR and subcloned into a pET28a vector between *EcoR* I and *Sal* I. The plasmid was amplified in DH5a and transformed into BL21(DE3) strain. C-terminal His-tagged GlnBP was expressed in *E. coli* BL21(DE3) in LB medium at 37 °C overnight with the induction of IPTG and purified by Ni²⁺-NTA agarose affinity chromatography. GuHCl was then added into GlnBP solution to a final concentration of 6 M to fully denature the protein and detach possibly existing L-glutamine. The mixture was purified by Superdex 75 size-exclusion chromatography column (GE Healthcare) to refold the protein and keep the purified ligand-free GlnBP in PBS buffer for further usage. In order to study the dynamics property of ligand-bound GlnBP, excess amount of L-glutamine was added into ligand-free GlnBP solution before each experiment.

Site-directed mutagenesis. The mutations of GlnBP in Thr118 were sequentially introduced by PCR reactions based on the plasmid *glnH*-pET28a as mentioned in the protein preparation part. In the first step, the DNA fragment of 5'-end to the mutation site and the fragment of mutation site to the 3'-end were synthesized by PCR separately. In the second step, the whole mutated *glnH* DNA was synthesized with the above two fragments as template. The mutated *glnH* DNA was then subcloned into the pET28a vector for plasmid amplification and protein purification. All plasmids were confirmed by sequencing.

Isothermal titration calorimetry. ITC measurements were performed on an ITC200 Micro calorimeter (MicroCal) at 18 °C. All samples were dissolved in 50 mM PBS buffer. The titrations were carried out by injecting 40- μ l aliquots of the ligand (0.5 and 1 mM ligand have used for GlnBP wild type and mutations, respectively) into protein (0.05 mM GlnBP wild type and 0.1 mM GlnBP

mutations, respectively) at time intervals of 2 min to ensure that the titration peak returned to the baseline. The titration data were analyzed using the program Origin7.0 and fitted with the one-site binding model.

smFRET experiment. The threonine 59 and threonine 130 of GlnBP were mutated to cysteines by standard PCR protocol. 50 μ M GlnBP-T59C/T130C was labeled by 200 μ M FRET donor (Alexa fluor 555-maleimide, Thermo Fisher Scientific Inc., MA, U.S.) and 400 μ M FRET acceptor (Alexa fluor 647-maleimide, Thermo Fisher Scientific Inc., MA, U.S.) by following the vendor provided protocol. The unreacted dye was separated from the labeled protein by using size-exclusion chromatography (SEC).

In order for immobilization of fluorescence-dye-labeled protein to carry out FRET experiment, the glass coverslip and drilled glass slide were cleaned by sonicating in water and ethanol three times respectively followed by being etched in plasma cleaner (PDC-002, Harrick Plasma Inc., NY, U.S.) for 5 min to destroy the residual dusts. The coverslip was stuck on the bottom of the drilled glass slide to make a flow cell. 100 μ l 0.1 mg/ml poly-lysine-PEG-NTA (PLL(20)-g[3.5]-PEG(2)-NTA, SuSoS AG Inc., Switzerland) solution was added to the flow cell and incubated for 20 min in order to create a layer of PEG on the coverslip surface and passivate it. The flow cell was washed with buffer thoroughly and incubated with 0.1 M NiCl₂ solution to introduce Ni²⁺ to the NTA. After 20 min incubation and complete wash, 1 nM fluorescence-dye-labeled GlnBP solution (50 mM phosphate buffer, pH 6.8) was added to the flow cell in order to tether the protein down to the glass surface upon the binding between the His-tag of GlnBP and the NTA group on the PEG layer.

The single molecule FRET images were taken by using a home-built wide-field fluorescence imaging system with an exposure time of 100 ms. The 532 nm laser beam was focused by a wide field 300 mm convex lens and then coupled to the microscope (IX-73, Olympus, Japan) through the back port. Eventually the beam illuminated a circular area on the coverslip and excited the labeled protein molecules in this area. The fluorescence of donor and acceptor was split by a dual image splitter (Optosplit II, Andor Technology Plc., U.K.) and the separated fluorescent images were detected by different areas of an EMCCD camera (iXon 897 Ultra, Andor Technology Plc., U.K.). The acceptor itself was excited at 632 nm in the control measurement.

The software iSMS⁹⁰ was used to extract and calculate the traces of fluorescence intensity of donor and acceptor from consecutively fluorescent images. We determined the number of FRET efficiency states by HMM analysis using vbFRET software⁹¹ (Fig. 6b). The change points of transitions between states in ideal HMM trajectories of ligand-bound GlnBP were used to build the transition density plot and it showed major transitions between four FRET efficiency states. A threshold analysis was performed to obtain the population of each state in both ligand-free and ligand-bound GlnBP systems. First, each point on TDP of ligand-bound GlnBP was dressed with a 2D normalized Gaussian function to facilitate its construction⁹². Then the constructed TDP was fitted by six 2D Gaussian functions to obtain the center and width of peaks. Finally, the population of each state in both ligand-free and ligand-bound GlnBP systems was obtained by using the threshold which set by the fullwidth half-height of the Gaussian distribution (Supplementary Fig. 5). These thresholds were ideal FRET 0.080–0.250 for E1, 0.255–0.415 for E2, 0.435–0.625 for E3 and 0.630–0.790 for E4. Histograms of FRET were obtained by combining all the time points of raw FRET trajectories corresponding to each state and fitted to the four Gaussian functions (Supplementary Fig. 6).

Statistics and reproducibility. Unless stated otherwise, the means and standard deviations for the quantities of each state were calculated based on all the MD conformations belonging to the specific state.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated and analyzed during this study are included in the article or from the corresponding author on request. Source data underlying plots shown in figures are provided in Supplementary Data 1.

Code availability

MSMbuilder3.8 was used for the MSM construction.

Received: 9 March 2020; Accepted: 14 July 2020;

Published online: 03 August 2020

References

- Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
- Clare, G. M. Interplay between conformational selection and induced fit in multidomain protein-ligand binding probed by paramagnetic relaxation enhancement. *Biophys. Chem.* **186**, 3–12 (2014).
- Greives, N. & Zhou, H. X. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proc. Natl Acad. Sci. USA* **111**, 10197–10202 (2014).
- Hammes, G. G., Chang, Y. C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA* **106**, 13737–13741 (2009).
- Zhou, H. X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* **98**, L15–L17 (2010).
- Chakrabarti, K. S. et al. Conformational selection in a protein-protein interaction revealed by dynamic pathway analysis. *Cell Rep.* **14**, 32–42 (2016).
- Nussinov, R., Ma, B. Y. & Tsai, C. J. Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys. Chem.* **186**, 22–30 (2014).
- Ma, B., Kumar, S., Tsai, C.-J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **12**, 713–720 (1999).
- Tsai, C.-J., Kumar, S., Ma, B. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **8**, 1181–1190 (1999).
- Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598 (1991).
- Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl Acad. Sci. USA* **44**, 98–104 (1958).
- Holyoak, T. In *Encyclopedia of Biophysics* (ed Roberts G.C.K.) 1584–1588 (Springer Berlin Heidelberg, 2013).
- Evenäs, J. et al. Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. *J. Mol. Biol.* **309**, 961–974 (2001).
- Vogt, A. D., Pozzi, N., Chen, Z. W. & Di Cera, E. Essential role of conformational selection in ligand binding. *Biophys. Chem.* **186**, 13–21 (2014).
- Michielssens, S. et al. A designed conformational shift to control protein binding specificity. *Angew. Chem. Int. Ed.* **53**, 10367–10371 (2014).
- Monod, J., Wyman, J. & Changeux, J. P. On nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–88 (1965).
- Moussatova, A., Kandt, C., O'Mara, M. L. & Tieleman, D. P. ATP-binding cassette transporters in *Escherichia coli*. *Biochim. Biophys. Acta* **1778**, 1757–1771 (2008).
- Dwyer, M. A. & Hellinga, H. W. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr. Opin. Struct. Biol.* **14**, 495–504 (2004).
- Hsiao, C. D., Sun, Y. J., Rose, J. & Wang, B. C. The crystal structure of glutamine-binding protein from *Escherichia coli*. *J. Mol. Biol.* **262**, 225–242 (1996).
- Sun, Y. J., Rose, J., Wang, B. C. & Hsiao, C. D. The structure of glutamine-binding protein complexed with glutamine at 1.94 angstrom resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* **278**, 219–229 (1998).
- Millet, O., Hudson, R. P. & Kay, L. E. The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy. *Proc. Natl Acad. Sci. USA* **100**, 12700–12705 (2003).
- Shen, Q., Simplaceanu, V., Cottam, P. F. & Ho, C. Proton nuclear magnetic resonance studies on glutamine-binding protein from *Escherichia coli*. Formation of intermolecular and intramolecular hydrogen bonds upon ligand binding. *J. Mol. Biol.* **210**, 849–857 (1989).
- Tang, C., Schwieters, C. D. & Clare, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
- D'Auria, S. et al. The differences in the microenvironment of the two tryptophan residues of the glutamine-binding protein from *Escherichia coli* shed light on the binding properties and the structural dynamics of the protein. *Proteins* **71**, 743–750 (2008).
- Bermejo, G. A., Strub, M. P., Ho, C. & Tjandra, N. Ligand-free open-closed transitions of periplasmic binding proteins: the case of glutamine-binding protein. *Biochemistry* **49**, 1893–1902 (2010).
- Pistolesi, S. & Tjandra, N. Temperature dependence of molecular interactions involved in defining stability of glutamine binding protein and its complex with L-glutamine. *Biochemistry* **51**, 643–652 (2012).
- Kim, E. et al. A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat. Chem. Biol.* **9**, 313–318 (2013).
- Liu, Z., Gong, Z., Guo, D. C., Zhang, W. P. & Tang, C. Subtle dynamics of holo glutamine binding protein revealed with a rigid paramagnetic probe. *Biochemistry* **53**, 1403–1409 (2014).
- Seo, M.-H., Park, J., Kim, E., Hohng, S. & Kim, H.-S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **5**, 3724 (2014).

30. Kooshapur, H., Ma, J., Tjandra, N. & Bermejo, G. A. The open conformation satisfies multiple NMR experiments on apo glutamine-binding protein: overcoming pitfalls in the study of interdomain dynamics. *Angew. Chem. Int. Ed.* **58**, 16899–16902 (2019).
31. Feng, Y. et al. Conformational dynamics of apo-GlnBP revealed by experimental and computational analysis. *Angew. Chem. Int. Ed.* **55**, 13990–13994 (2016).
32. Flocco, M. M. & Mowbray, S. L. The 1.9 Å x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Biol. Chem.* **269**, 8931–8936 (1994).
33. Oh, B. H., Ames, G. F. & Kim, S. H. Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *J. Biol. Chem.* **269**, 26323–26330 (1994).
34. Oswald, C. et al. Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J. Biol. Chem.* **283**, 32848–32859 (2008).
35. Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657–10663 (1992).
36. Spurlino, J. C., Lu, G. Y. & Quiocho, F. A. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266**, 5202–5219 (1991).
37. Quiocho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**, 17–25 (1996).
38. Bowman, G. R., Voelz, V. A. & Pande, V. S. Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* **21**, 4–11 (2011).
39. Chodera, J. D. & Noe, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).
40. Huang, X., Bowman, G. R., Bacallado, S. & Pande, V. S. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl Acad. Sci. USA* **106**, 19765–19769 (2009).
41. Prinz, J. H. et al. Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
42. Bowman, G. R., Huang, X. & Pande, V. S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197–201 (2009).
43. Nuske, F., Keller, B. G., Perez-Hernandez, G., Mey, A. S. J. S. & Noe, F. Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **10**, 1739–1752 (2014).
44. Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noe, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
45. Schwantes, C. R. & Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 (2013).
46. Harrigan, M. P. et al. MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* **112**, 10–15 (2017).
47. Long, C. H., Chao, E., Da, L. T. & Yu, J. Determining selection free energetics from nucleotide pre-insertion to insertion in viral T7 RNA polymerase transcription fidelity control. *Nucleic Acids Res.* **47**, 4721–4735 (2019).
48. Zhang, L. et al. Elucidation of the dynamics of transcription elongation by RNA Polymerase II using kinetic network models. *Acc. Chem. Res.* **49**, 687–694 (2016).
49. Noe, F. & Rosta, E. Markov models of molecular kinetics. *J. Chem. Phys.* **151**, 190401 (2019).
50. Husic, B. E. & Pande, V. S. Markov state models: from an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396 (2018).
51. Olsson, S., Wu, H., Paul, F., Clementi, C. & Noe, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl Acad. Sci. USA* **114**, 8265–8270 (2017).
52. Wang, W., Cao, S., Zhu, L. & Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, 1–18 (2018).
53. Keller, B. G., Kobitski, A., Jaeschke, A., Nienhaus, G. U. & Noe, F. Complex RNA folding kinetics revealed by single-molecule FRET and hidden Markov models. *J. Am. Chem. Soc.* **136**, 4534–4543 (2014).
54. Bucher, D., Grant, B. J., Markwick, P. R. & McCammon, J. A. Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS Comput. Biol.* **7**, e1002034 (2011).
55. Bucher, D., Grant, B. J. & McCammon, J. A. Induced fit or conformational selection? The role of the semi-closed state in the maltose binding protein. *Biochemistry* **50**, 10530–10539 (2011).
56. Gu, S., Silva, D.-A., Meng, L., Yue, A. & Huang, X. Quantitatively characterizing the ligand binding mechanisms of choline binding protein using Markov state model analysis. *PLoS Comput. Biol.* **10**, e1003767 (2014).
57. Silva, D.-A., Bowman, G. R., Sosa-Peinado, A. & Huang, X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput. Biol.* **7**, e1002054 (2011).
58. Wang, Y., Tang, C., Wang, E. K. & Wang, J. Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein. *PLoS Comput. Biol.* **8**, e1002471 (2012).
59. Su, J. et al. Analysis of domain movements in glutamine-binding protein with simple models. *Biophys. J.* **92**, 1326–1335 (2007).
60. Kumari, R., Kumar, R., Lynn, A. & Consortium, O. S. D. D. g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Modeling* **54**, 1951–1962 (2014).
61. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037 (2001).
62. Wang, E. et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* **119**, 9478–9508 (2019).
63. Wang, C., Greene, D. A., Xiao, L., Qi, R. & Luo, R. Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* **4**, 87 (2018).
64. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA* **106**, 13737 (2009).
65. Cserehely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35**, 539–546 (2010).
66. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507 (2008).
67. Ha, T. et al. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264 (1996).
68. Sun, H., Li, Y., Tian, S., Xu, L. & Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **16**, 16719–16729 (2014).
69. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
70. Sorin, E. J. & Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **88**, 2472–2493 (2005).
71. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
72. Wang, J. M., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
73. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
74. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
75. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
76. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. In *Solving Software Challenges for Exascale*. (eds Markidis, S. & Laure, E.) 3–27 (Springer International Publishing, 2015).
77. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
78. Sheong, F. K., Silva, D.-A., Meng, L., Zhao, Y. & Huang, X. Automatic state partitioning for multibody systems (APM): an efficient algorithm for constructing Markov state models to elucidate conformational dynamics of multibody systems. *J. Chem. Theory Comput.* **11**, 17–27 (2015).
79. Zhao, Y., Sheong, F. K., Sun, J., Sander, P. & Huang, X. A fast parallel clustering algorithm for molecular simulation trajectories. *J. Comput. Chem.* **34**, 95–104 (2013).
80. Bowman, G. R., Beauchamp, K. A., Boxer, G. & Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* **131**, 124101 (2009).
81. Buchete, N. V. & Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **112**, 6057–6069 (2008).
82. Malmstrom, R. D., Lee, C. T., Van Wart, A. T. & Amaro, R. E. Application of molecular-dynamics based Markov state models to functional proteins. *J. Chem. Theory Comput.* **10**, 2648–2657 (2014).
83. Pan, A. C. & Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 064107 (2008).

84. Yao, Y. et al. Hierarchical Nystrom methods for constructing Markov state models for conformational dynamics. *J. Chem. Phys.* **138**, 174106 (2013).
85. McGibbon, R. T. & Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **142**, 124105 (2015).
86. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. & Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **126**, 155101 (2007).
87. Swope, W. C., Pitera, J. W. & Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. I. Theory. *J. Phys. Chem. B* **108**, 6571–6581 (2004).
88. Deuffhard, P., Huisinga, W., Fischer, A. & Schutte, C. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl.* **315**, 39–59 (2000).
89. Deuffhard, P. & Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **398**, 161–184 (2005).
90. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. iSMS: single-molecule FRET microscopy software. *Nat. Methods* **12**, 593–594 (2015).
91. Bronson, J. E., Fei, J. Y., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys. J.* **97**, 3196–3205 (2009).
92. Pirchi, M. et al. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* **2**, 493 (2011).

Acknowledgements

We thank the National Natural Science Foundation of China Grants (21733007, 21803071 (to L.Z.) and 21773038 (to W.W.)), the National 1000 Youth Talents Program of China (to L.Z.), and the Ministry of Science and Technology of the People's Republic of China (2016YFA0501702 (to W.W.)) for the financial support. We acknowledge the insightful discussions with Prof. Xuhui Huang and Dr. Siqin Cao.

Author contributions

L.Z. and W.W. conceived and designed the project. Y.F., S.W., and D.W. performed the biochemical and spectroscopic experiments. Y.F., S.W., Z.L., J.L., and D.W. analyzed the experimental results. L.Z. performed computational studies. L.Z., Y.F., S.W., X.J., and W.W. prepared the manuscript. All authors contributed to data interpretation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-01149-z>.

Correspondence and requests for materials should be addressed to L.Z. or W.W.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020