PCCP

PAPER

Exploration of gated ligand binding recognizes an allosteric site for blocking FABP4–protein interaction†

Yan Li,‡*[a] Xiang Li‡[b] and Zigang Dong*[a]

Fatty acid binding protein 4 (FABP4), reversibly binding to fatty acids and other lipids with high affinities, is a potential target for treatment of cancers. The binding site of FABP4 is buried in an interior cavity and thereby ligand binding/unbinding is coupled with opening/closing of FABP4. It is a difficult task both experimentally and computationally to illuminate the entry or exit pathway, especially with the conformational gating. In this report we combine extensive computer simulations, clustering analysis, and the Markov state model to investigate the binding mechanism of FABP4 and troglitazone. Our simulations capture spontaneous binding and unbinding events as well as the conformational transition of FABP4 between the open and closed states. An allosteric binding site on the protein surface is recognized for the development of novel FABP4 inhibitors. The binding affinity is calculated and compared with the experimental value. The kinetic analysis suggests that ligand residence on the protein surface may delay the binding process. Overall, our results provide a comprehensive picture of ligand diffusion on the protein surface, ligand migration into the buried cavity, and the conformational change of FABP4 at an atomic level.

[a] *The Hormel Institute, University of Minnesota, Austin Minnesota 55912, USA.
E-mail: yanli0208@hi.umn.edu, zgdong@hi.umn.edu; Fax: +1-(507)-437-9606;
Tel: +1-(507)-437-9600*

[b] *Department of Physiology and Pathophysiology, School of Basic Medical Sciences,
Zhengzhou University, 450001 Zhengzhou, Henan, China*

† Electronic supplementary information (ESI) available: Figures showing additional MD information, the hierarchical clustering results, additional 1D and 2D FESs, FABP4 conformational distribution as a function of the ligand RMSD, calculation of the bound volume, and evolution of the implied timescale, and tables listing the net flux, MFPT, association and dissociation rate constants, and calculated binding affinities. See DOI: 10.1039/c5cp04784f
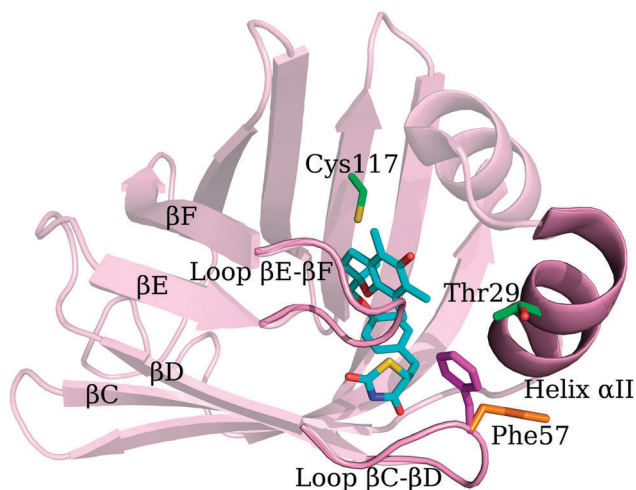
‡ Y.L. and X.L. contributed equally to this work.

Introduction

Fatty acid binding proteins (FABPs), mediating transport, storage and metabolism of fatty acids and other lipids, escort and sequester ligands to regulate signaling pathways and enzyme activities.[1] Among nine FABP family members, FABP4 has been recognized as a potential target for treatment of type 2 diabetes, atherosclerosis, and ovarian cancer.[2,3] The structural feature of FABP4 has been well studied and high-resolution crystal structures of both the open and closed states have been solved. However, the mechanism of ligand binding/unbinding that is the molecular basis for FABP4's functions remain to be clarified. The ligand-binding site of FABP4 is buried in an interior cavity and surrounded by two α-helices and ten β-strands,

which are shared by other member proteins. The portal region, including helix αII and loops between βC–βD and βE–βF (Fig. 1), has long been postulated to be the entrance into the binding cavity.[4–6] Phe57 at the mouth of the portal is regarded as the gating residue. In crystal structures, the side chains of Phe57 adopt different conformations in the open and closed states (Fig. 1). According to the classification of enzyme gates,[7] FABP4 has a wing-type gate. Previously we have demonstrated the conformational transition of FABP4 between the open and closed forms in the absence and presence of different ligands.[8] However, the coupling of FABP4's conformational changes and ligand binding has yet not been elucidated. A detailed description of FABP4–ligand binding will be helpful in the development of potent FABP4 inhibitors for treatment of diabetes and cancers.

Molecular dynamics (MD) simulations have already been utilized to provide insights into the mechanism of ligand binding and protein conformational changes. The dynamic description of gated binding dates back to more than thirty years ago with work on migration pathways in myoglobin.[9] Since then, the gating model has been built and its contribution to the binding rate has been evaluated.[10–12] So far diffusion of gaseous molecules (CO, CO_2, O_2, and NO) within a gated channel has been widely studied in several proteins.[13–21] And the computer simulations have been compared with experimental measurements.[22,23] In recent years, the binding/unbinding process has been

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 32257–32267 | **32257**

Fig. 1 The open and closed states of FABP4. Two FABP4 structures (PDB entries 2QM9 and 3HK1) are superimposed using backbone atoms. In the occluded state (2QM9), Phe57 (magenta) points towards Thr29 and blocks access to the cavity. In the open state (3HK1), Phe57 (orange) projects outwardly away from Thr29 and exposes the ligand to the solvent. The portal, including helix αII and loops between βC–βD and βE–βF, is highlighted. Cys117 is located at the bottom of the binding cavity. Thr29, sitting on helix αII, and Phe57, sitting on loop βC–βD, are located at the mouth of the portal. The ligand with carbon atoms colored cyan indicates the binding cavity. Carbon atoms of Thr29 and Cys117 are colored green. The ligand and residues are shown in sticks.

investigated[24–27] whereas the conformational gating is not considered. Ligand diffusion on the protein surface has been reported as well.[28,29] However, the mechanism of drug-like small molecule diffusion into a buried cavity has been very rarely reported. It still remains a challenge to achieve a comprehensive picture of gated ligand binding because the binding site is enclosed and the coupling of conformational gating and ligand diffusion needs to be adequately sampled.

To capture the underlying mechanism of FABP4–ligand binding, we extend the pathway sampling strategy[30] recently reported to investigate the diffusion of troglitazone (TGZ, Fig. 2), a drug-like small molecule inhibitor, into FABP4. In order to sample the conformational space more efficiently, two hundred seeding structures are extracted along an unbinding pathway generated by steered MD[31] (SMD). Then each of them is subject to a 200 ns MD simulation and the ensemble of MD simulations
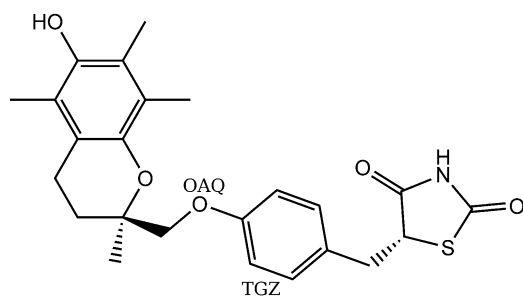

Fig. 2 Chemical structure of troglitazone (TGZ). The oxygen atom of OAQ is labeled.

is labeled as MD200. Clustering analysis of MD200 reveals the significant regions on the free energy landscape. Furthermore, to avoid possible biases introduced by the SMD simulation, two hundred and forty structures are obtained from the important regions, and a 400 ns MD simulation is initiated on each of them. This ensemble of MD simulations is labeled as MD240. Clustering analysis and the Markov state model (MSM) are employed to construct the binding pathway, which provides novel insights into the mechanism of gated ligand binding. We demonstrate that the opening/closing of FABP4 and ligand diffusion are well sampled. A binding pocket on the protein surface is identified to be the target site for the development of novel FABP4 inhibitors. Our results provide thermodynamic and kinetic details of FABP4–TGZ binding, including ligand diffusion on the protein surface, ligand penetration into the internal cavity, and thermal fluctuations of FABP4.

Results

Spontaneous binding/unbinding events

In accordance with previous experimental reports,[4–6] the ligand is pulled out of the binding cavity through the portal in the SMD simulation. The movement of TGZ is gauged by the heavy-atom root mean square deviation (RMSD) of the ligand with respect to the crystal structure as well as the distance between centers of mass (COMs) of TGZ and Cys117/Thr29. The openness of the portal is evaluated by the distance between COMs of Thr29 and Phe57. The global motion of FABP4 is measured by the RMSD of backbone heavy atoms. In the first 0.4 ns, the Thr29–ligand distance decreases and the Cys117–ligand distance increases (Fig. 3), indicating that TGZ moves close to the portal. After that, the distance both increases, suggesting that the ligand floats away from the protein. At around 0.4 ns, a peak appears in the curve of the Thr29–Phe57 distance suggesting the conformational transition of FABP4 from the closed form to the open form. At the same time the backbone RMSD has a slight change (<1 Å). The results suggest that the conformational change of Phe57 is needed to facilitate ligand unbinding while a subtle change of the backbone conformation is sufficient for TGZ to escape from the internal cavity.

Two hundred seeding structures extracted from the unbinding pathway work as the starting structures, and each one is subject to a 200 ns MD simulation. According to the initial position of the ligand (ESI,† Table S1), the 200 MD trajectories (MD200) are grouped into three ensembles. Ensemble 1 (E1) comprises the first 71 trajectories started with the ligand in the interior cavity. Ensemble 2 (E2) is composed of the following 44 trajectories (Traj 72–115) begun with the ligand around the portal. Ensemble 3 (E3) consists of the last 85 trajectories (Traj 116–200) started with the ligand in the exterior solvent. The poses of the ligand at various time points are displayed in Fig. 3.

For E1, only in trajectory 16 the unbinding event is observed with the maximal ligand RMSD over 20 Å (Fig. 4A). Inspection of the trajectory reveals that the ligand escapes through the portal, adheres to the protein surface, and finally re-enters the

32258 | *Phys. Chem. Chem. Phys.*, 2015, **17**, 32257–32267

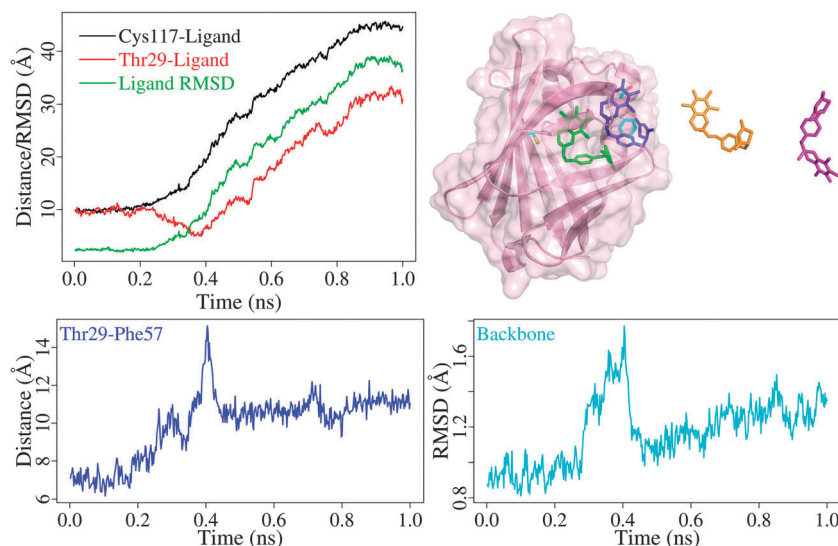This journal is © the Owner Societies 2015

Fig. 3 Pulling TGZ out of the binding cavity in the SMD simulation. Time evolution of the Cys117–ligand distance (black line), the Thr29–ligand distance (red line), the ligand RMSD (green line), the Thr29–Phe57 distance (blue line), and the backbone RMSD (cyan line) is plotted. Snapshot structures of the ligand at 0 ps, 360 ps, 580 ps, and 1000 ps are shown in green, blue, orange, and magenta, respectively. Carbons of three residues (Thr29, Phe57 and Cys117) are colored cyan. At 360 ps (time frame 72), TGZ is in contact with water molecules in the exterior solvent (within 5 Å of the ligand). At 580 ps (time frame 116), the ligand is completely out of contact with the protein when none of protein atoms is within 5 Å of the ligand.
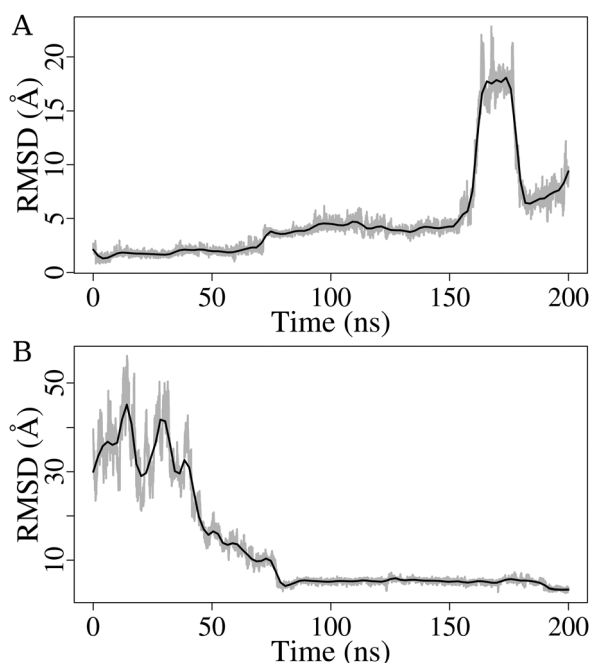

Fig. 4 Time evolution of the ligand RMSD in the trajectories 16 (A) and 178 (B). RMSD values of sampled structures are shown in gray. The smoothed line, computed with LOESS implemented in the statistical software R v3.0, is plotted in black.

pocket through the portal again. For E2, in 10 out of 44 trajectories the minimal ligand RMSD is below 3 Å, suggesting that TGZ successfully reaches the binding pose experimentally determined. For E3, in 30 out of the 85 trajectories the minimal RMSD is less than 10 Å suggesting that the ligand successfully penetrates into the cavity. In two trajectories the minimal RMSD is below 3 Å, suggesting that a full binding pathway

is achieved. In one pathway (Traj 178, Fig. 4B), the ligand keeps jumping between the protein surface and the solvent in the first 40 ns, exploring the whole protein surface, then attaches to the portal and lingers on it for 20 ns, and finally finds the tunnel into the binding cavity at 70 ns. Similar to the SMD simulation, when the ligand moves into or out of the protein, FABP4 is in the open state and the backbone RMSD has no notable change (ESI,† Fig. S1).

In this work the ligand RMSD of 10 Å is chosen as the threshold value to judge if the ligand enters or leaves the internal cavity. If the RMSD is below 10 Å, the ligand is inside the binding cavity; otherwise, the ligand is outside of the protein. This value is selected based on the analysis of MD trajectories and the results mentioned later. In the first 81 MD simulations, the starting structure has a ligand RMSD under 10 Å. Only in one trajectory (Traj 16), the ligand temporarily escapes from the binding cavity. We postulate that unbinding occurs scarcely after the ligand enters the buried cavity. Thus, the ligand RMSD of 10 Å is determined as the criterion. The following results also support that this threshold value is reasonable. Similarly, the RMSD is used to be indicative of the water diffusion in a recent report.[32]

Transition states of the ligand

Clustering analysis is employed to identify metastable states of the ligand during the binding process. Here we focus on ligand diffusion and thereby the conformational change of the protein and the ligand is ignored in clustering. The conformational change of FABP4 will undoubtedly exert influence on ligand diffusion and *vice versa*. The effect should be reflected in ligand distribution and will be discussed later. The oxygen atom of OAQ at the center of the TGZ structure (Fig. 2) is chosen to represent the ligand molecule. In our opinion it is simple but

good enough to describe the ligand movement with single atom's coordinates. The clustering is performed in the three-dimensional space of the OAQ's coordinates using Bayesian clustering, which has been employed for analysis of MD trajectories.[30,33] In the optimal resolution 77 classes (class 0–76) are obtained and a representative structure is selected for each class.

In six classes (class 27, 31, 52, 56, 59, and 68), TGZ is in the exterior solvent with the average ligand RMSD over 30 Å (Table 1). Visual examination shows that these classes are distributed in distinct directions around the protein (Fig. 5A). Therefore, the six classes are considered to be the unbound state. The other 71 classes are then clustered into six macrostates (C1–C6) by hierarchical clustering (ESI,† Fig. S2). C1 is located inside the binding cavity with the average RMSD of 4.2 Å (Table 1) and thereby is the bound state. Five populated sites (C2–C6) on the protein surface are identified. Among them, C2 is situated on the portal with the average RMSD of 13.0 Å. These results suggest that the ligand RMSD of 10 Å is an appropriate criterion to assess if the ligand is inside the interior cavity. The representative structures for C1–C6 are displayed in Fig. 5B. The unbound

state is labeled as C7. Thus, the 77 classes generated by Bayesian clustering are regarded as microstates and further grouped into seven macrostates.

The coordinates of the OAQ atom are projected onto the two-dimensional plane (xy, xz, and yz), and the free energy surface (FES) is calculated (Fig. 5C and Fig. S3, ESI†). According to the clustering results and FES, the protein surface and internal cavity are well explored by the ligand. The six macrostates (C1–C6) are directly associated with energy wells on the FES. The convergence of free energy calculation is shown in ESI,† Fig. S4.

Binding kinetics

In the dataset MD200, the starting structures are extracted from a SMD trajectory generated with a biasing potential. Although it looks that the protein surface is well explored, our results may be compromised due to the biases introduced in the SMD simulation. In order to fully explore the entire potential energy landscape and avoid possible biases, another ensemble of starting structures is acquired from the significant regions indicated by the FES and clustering analysis.

Table 1 Locations and mean values of the ligand RMSD with standard deviations for the macrostates C1–C7

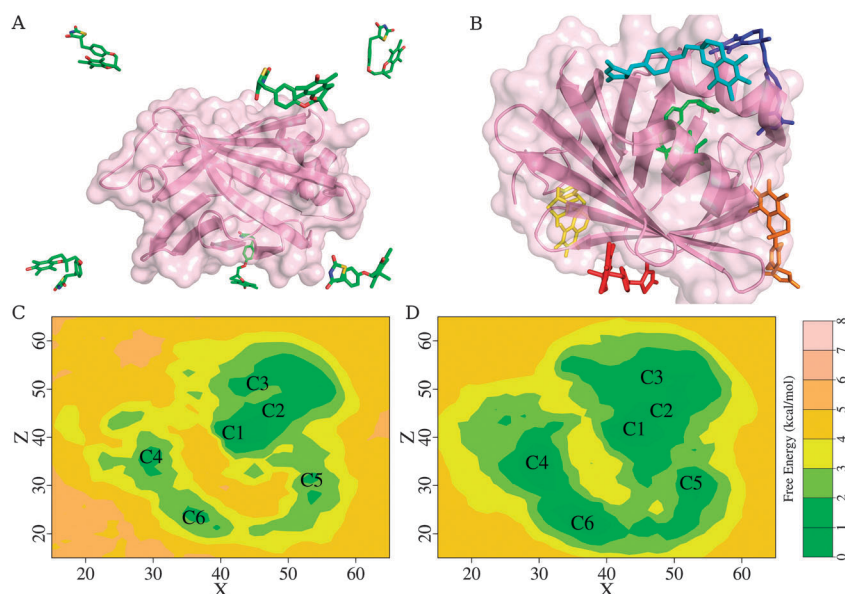| State	MD200		MD240		Location
	Number of classes	RMSD (Å)	Number of classes	RMSD (Å)	
C1	23	4.2 ± 1.8	7	6.7 ± 2.1	Inside the binding cavity
C2	19	13.0 ± 3.8	12	13.6 ± 2.9	Around the portal
C3	11	16.6 ± 2.2	7	18.6 ± 3.8	Between helix αI and αII
C4	8	22.9 ± 2.4	8	22.0 ± 2.6	N terminus on the backside
C5	4	19.0 ± 2.2	10	21.5 ± 1.9	Among loops αI–αII/βE–βF/βG–βH
C6	6	20.5 ± 1.8	11	20.3 ± 1.6	Strands βH/βI/βJ
C7	6	32.0 ± 9.2	4	36.1 ± 7.1	In the exterior solvent

Fig. 5 Transition states of the ligand and the 2D FES. (A) The unbound state C7. The representative structures for 6 classes which comprise the macrostate C7 are displayed. (B) The six states (C1–C6) on the protein surface or inside the protein. The representative structure for C1–C6 is colored green, blue, cyan, yellow, orange, and red, respectively. (C) The 2D FES obtained from MD200. (D) The 2D FES obtained from MD240. Points sampled from MD simulations are assigned into 75 × 75 bins for MD200 and 75 × 80 bins for MD240 on the xz plane with a bin area of 1.4 × 1.4 Å². The macrostates (C1–C6) on the FES correspond to the binding sites in the panel (B).

32260 | *Phys. Chem. Chem. Phys.*, 2015, **17**, 32257–32267

This journal is © the Owner Societies 2015

The six macrostates (C1–C6) as well as the six microstates in C7 are grouped into 20 clusters, respectively, using the k-means method. The centroid structure in each of the 240 clusters works as the starting point for a 400 ns MD simulation (MD240). In 20 trajectories started inside the protein (C1), none of escaping events is observed. The other 220 MD simulations are initiated with the ligand outside of the protein. In 82 out of 220 trajectories the minimal RMSD is below 10 Å; in 66 instances the final RMSD, which is computed as the average of the ligand RMSD in the last 1 ns, is less than 10 Å; and in 10 trajectories the minimal RMSD is less than 3 Å. Thus, in more than one third of MD simulations (82/220) the binding events have occurred; in most of them (66/82) the ligand finally stays inside the protein; and in 10 trajectories the ligand successfully reaches the crystal position.

Similarly, the Bayesian clustering has been performed in the three-dimensional space of the atom OAQ's coordinates. In the optimal resolution 59 classes (0–58) are obtained. Among them, four classes (43, 45, 47, and 51) are far away from the protein with an average RMSD over 30 Å (Table 1), which is the unbound state (C7). Other 55 classes are grouped into six macrostates (C1–C6) using the hierarchical clustering method (ESI,† Fig. S5). Similar to Fig. 5A, in C7 the ligand is distributed in four distinct directions. In C1–C6, the ligand is positioned at the same site as shown in Fig. 5B. The FES on the two-dimensional plane is displayed in Fig. 5D and Fig. S6 (ESI†), and the convergence of the free energy calculation is shown in ESI,† Fig. S7.

Transitions between these states are analyzed by MSM, which has been widely employed in studies of the protein conformational change and protein–ligand interaction.[24,34–39] In this work, the seven macrostates are used to construct the transition matrix $T(\tau)$. The lag time (τ) is determined to be 80 ns after analysis of the implied timescale (ESI,† Fig. S8). The binding kinetics between the unbound state (C7) and the bound state (C1) is described with MSM and transition path theory (TPT). The binding time is estimated by the mean first passage time (MFPT) to the final state (C1) from the other states.

As revealed in the previous report,[8] the ligand is highly mobile inside the binding cavity and there may exist other binding modes except the crystal one. It explains the observation that binding events occur in 82 trajectories while only in 10 instances the ligand reaches the crystal position. For MSM calculation, it is important to obtain enough transitions between macrostates such as the transition between C1 and C2. After the ligand enters the internal cavity, the diffusion to the crystal position occurs within C1, which makes little contribution to the MSM calculation. Therefore, it is not necessary to have many full binding trajectories. With binding events observed in more than one third of MD simulations, our data are good enough for the MSM calculation.

The total net flux between C1 and C7 is $(4.3 \pm 0.2) \times 10^4 \text{ s}^{-1}$. The net flux between pairs of states is listed in ESI,† Tables S2 and S3. For clarity of visualization, the flux network shown in Fig. 6 is constituted by the transitions with the flux above $1.0 \times 10^3 \text{ s}^{-1}$. We notice that C2 plays a vital role during the binding process. The flux from C2 to C1, $(1.2 \pm 0.1) \times 10^4 \text{ s}^{-1}$, accounts for $(28.5 \pm 0.6)\%$ of the total flux. The influx of C1, $(1.6 \pm 0.1) \times 10^4 \text{ s}^{-1}$, occupies $(36.4 \pm 0.7)\%$ of the total flux, 75% of which comes from C2, suggesting that C2 acts as an indispensable intermediate to the bound state. The direct transition to C1 from other states except C2, accounting for 7.9% of the total flux, indicates a fast binding event in which the ligand passes through the portal quickly as shown in Fig. 4B.

The binding time from the solvent (C7) to the internal cavity (C1) is 1.64 ± 0.11 µs (ESI,† Table S4). Ligand residence on C4–C6 delays the binding by increasing the MFPT to at least 1.88 ± 0.12 µs. Furthermore, we calculate the binding time in 120 simulations starting with the ligand in the solvent, far away from the protein. The MFPT time is 1.28 ± 0.06 µs (ESI,† Table S5). The transition time from C2 to C1 is 0.97 ± 0.07 µs, which occupies about 60% of the binding time, suggesting that ligand diffusion through the portal is the rate-limiting step during the binding process. It is consistent with the mutation and kinetics studies.[4,6] Examination of binding trajectories show that the ligand attaches to and detaches from the protein surface very fast, hops between different surface pockets, and attempts to pass through the portal, but penetrates into the protein slowly. A traditional view about protein–ligand binding is that the ligand nonspecifically binds with the protein and subsequently diffuses on the protein surface, which speeds the search for the
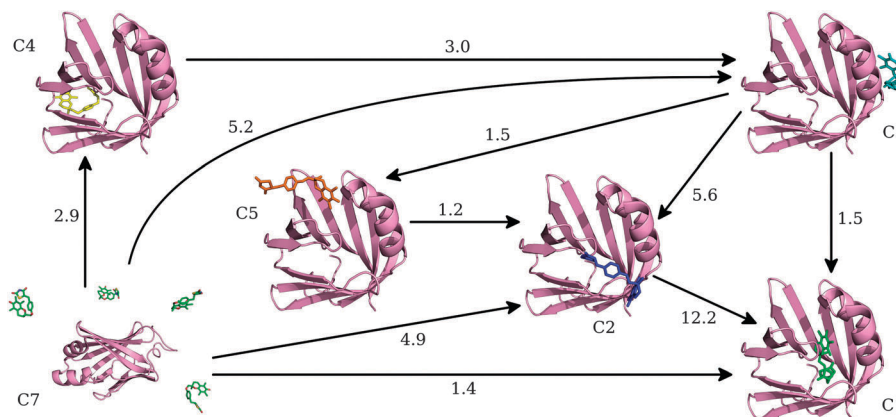

Fig. 6 Flux network of ligand binding from the unbound state C7 to the bound state C1. The representative structure for each state is displayed. The net flux (ms^{-1}) is labeled next to the arrow.

binding site.[40] Our results suggest that ligand residence on the protein surface may not necessarily accelerate the binding. The ligand may be trapped in some surface pockets, which delays the transition into the binding cavity.

An allosteric site for FABP4–protein interaction

Examination of the populated sites on the protein surface reveals three binding pockets located in C3, C4, and C5, respectively. Among them, the binding pocket in C3, lying between helices αI and αII, may be involved in FABP4–protein interactions. FABP4 can associate with another protein and plays a regulatory role as a lipid sensor.[41] Although it is still unclear what the FABP4–protein complex looks like, the mutation study has demonstrated that four charged residues (Asp17, Asp18, Lys21, and Arg30) are required in the interaction between FABP4 and hormone sensitive lipase (HSL).[42] Arg30 is situated on helix αII and the other three residues are seated on helix αI (Fig. 7). In crystal structures, Asp17 and Arg30 form an ion bond. The distance between OD1/OD2 of Asp17 and NH2 of Arg30 is around 4.1 Å in the structure of 2QM9, which makes this site unsuitable for ligand binding. In the simulations, due to conformational changes, the ion pair Asp17–Arg30 is broken and an orifice appears between helices αI and αII, which make enough room to accommodate the ligand. This pocket can be a target site for discovery of novel FABP4 inhibitors to mediate the protein function by blocking the FABP4–protein interaction.

Opening and closing of FABP4

As demonstrated in the previous report,[8] FABP4 exists in equilibrium between the open and closed form, and opening and closing of FABP4 are influenced by the bound ligand. Here we investigate the conformational change of FABP4 between the open and closed state during ligand binding as well as its effect on the binding rate. The conformational distribution of FABP4 as a function of the Thr29–Phe57 distance is shown in Fig. 8A. In all density curves, two population peaks at around 7.0 and 10.0 Å are observable, which correspond to the closed and open form, respectively. The results also suggest that opening and closing of FABP4 are well sampled in our simulations.

The population of the closed form (Fig. 8B) is then acquired as the integral of the density curve when the ligand is situated in different locations. The closed state is defined to be all conformations with the distance between Thr29 and Phe57 less than 8.5 Å.[8] When the ligand fluctuates inside the protein (C1), the population of the closed form is 29% and thereby the open conformation is preferred, which is consistent with the previous report.[8] Ligand residence on C2 results in opening of the portal with the lowest population (22%) of the closed conformation, which facilitates ligand entry into the protein. An orifice appears on the surface site C3, C4, and C5 when the ligand is located there. Opening of C3, C4, and C5 leads to closure of C2 because of relative movements of the secondary structures, which explains the rising population of the occluded form (47–57%). By contrast, the closed form constitutes 62% of the total population in the absence of a ligand.[8] Moreover, we investigate the conformational distribution of FABP4 as a function of the ligand RMSD (ESI,† Fig. S9) as
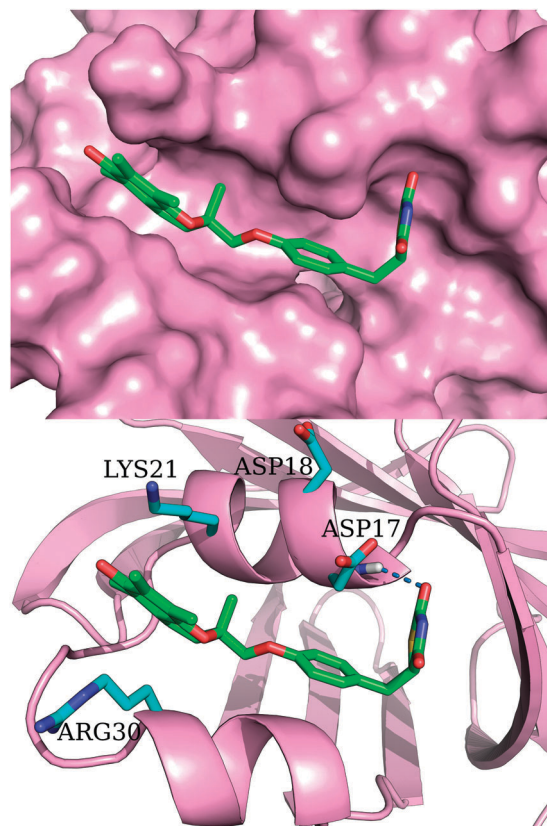

Fig. 7 The binding mode of TGZ in the surface pocket of C3. FABP4 is shown in the surface and cartoon, respectively. In the surface representation, a small orifice is observable in the binding pocket. The four residues (Asp17, Asp18, Lys21 and Arg30) are shown in sticks with carbon atoms colored cyan. The hydrogen bond between TGZ and Asp17 is shown in the dashed line. TGZ is shown in sticks with carbon atoms colored green.

well as the distribution in the dataset MD200 (ESI,† Fig. S10). The results agree very well with the data shown in Fig. 8 and Table 1.

A model depicting the kinetics for protein conformational changes has been constructed.[43,44] Here, the conformational transition of FABP4 is described by a simple two-state MSM having the open and closed state. The constant rates for opening and closing are computed to be $k_{open} = (5.7 \pm 0.05) \times 10^6 \text{ s}^{-1}$ and $k_{close} = (3.2 \pm 0.06) \times 10^6 \text{ s}^{-1}$ (ESI,† Table S6), respectively. The mean times for opening and closing are $\text{MFPT}_{open} = 175.3 \pm 1.5$ ns and $\text{MFPT}_{close} = 309.8 \pm 5.4$ ns. Based on the gating model,[12] the effect of the conformational gating on ligand binding can be estimated as $k_{on}/k_{on}' \approx k_{open}/(k_{open} + k_{close}) = 0.64$, where k_{on} is the rate constant for gated ligand binding and k_{on}' is the rate constant in the absence of gating. Thus, due to hindrance of the conformational gating, the binding rate decreases about 40%.

Estimation of binding affinity

In this section, the binding affinity of FABP4 and TGZ is calculated first based on the MSM model, and then based on the FES obtained from both MD200 and MD240. The experimental K_d value of FABP4–TGZ is 17.0 nM[45] and thereby the binding affinity is $\Delta G = RT \ln K_d = -10.66$ kcal mol^{-1}. With the seven-state model, the association constant is computed to be

32262 | *Phys. Chem. Chem. Phys.*, 2015, **17**, 32257–32267

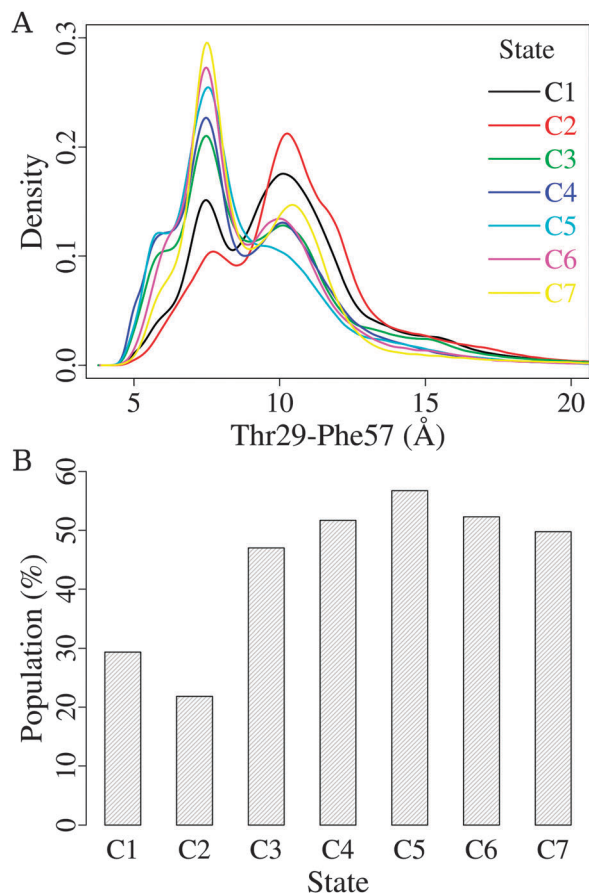This journal is © the Owner Societies 2015

Fig. 8 Conformational distribution of FABP4 between the open and closed form. (A) Distribution of FABP4 conformations in the seven macrostates. (B) Population of the closed form. The density curves are calculated using R v3.0. The opening/closing of FABP4 is described by the distance between Thr29 and Phe57. The closed form is defined to be all conformations with the Thr29–Phe57 distance below 8.5 Å.

$k_{on} = 1.51 \times 10^8$ M^{-1} s^{-1} and the dissociation constant is $k_{off} = 4.60 \times 10^4$ s^{-1} (see Methods and ESI,† Table S7). Thus, the binding affinity is $\Delta G = RT \ln(k_{off}/k_{on}) = -4.8$ kcal mol^{-1}, which is only half of the experimental value.

Methods for the calculation of protein–ligand binding affinity from MD simulations have been discussed in depth.[46] In this work, the binding affinity of FABP4–TGZ is estimated using the protocol previously reported.[47] To do it, the three-dimensional FES is constructed after assigning the coordinates of the atom OAQ into bins (see Methods). The bound volume (V_b) is required for calculation of the binding affinity. The sensitivity analysis (ESI,† Fig. S11) yields a bound volume of 675.37 Å^3 for the dataset MD200 and 860.34 Å^3 for the dataset MD240 with the free energy no more than 4.0 kcal mol^{-1}. Thus, the binding affinity is estimated to be -7.86 kcal mol^{-1} and -7.96 ± 0.07 kcal mol^{-1} (ESI,† Table S8), respectively. The difference between the calculation and the experiment is less than 3.0 kcal mol^{-1}.

With MD200, we investigate the influence of the number of MD trajectories on the calculation. Three subsets are built with 100, 140, and 180 MD trajectories, which are randomly selected from ensembles E1, E2, and E3 (ESI,† Table S9). The binding affinity

acquired is -7.42, -7.63, and -7.79 kcal mol^{-1}, respectively. When the number of trajectories rises from 100 to 200, the change of estimated binding affinity is less than 0.5 kcal mol^{-1}.

Conclusion

In this report we demonstrate how to build the unbiased distribution describing ligand binding on the basis of a biased path. The dataset MD200 is obtained by performing additional MD simulations at given points along the pathway defined by the SMD simulation and thereby is somewhat biased. The dataset MD240 is initialized with points uniformly sampled around the protein and is supposed to be unbiased. Interestingly, the binding affinities computed from the two datasets are very close though MD240 has more starting structures and longer simulations than MD200. Examination of FESs (1D and 2D) implies that the protein surface is better sampled in MD240 than in MD200 whereas it hardly improves the prediction of the binding affinity. Even compared to the subsets of MD200, MD240 does not lead to a notably better prediction. Moreover, the MSM calculation is also performed with the dataset MD200 (ESI,† Fig. S12 and Tables S10–S11). The conclusions that can be drawn from MD200 are almost the same with those from MD240. This raises a series of intriguing questions. Is it necessary to equally explore the full conformational space for thermodynamic and kinetic analysis? If not, which part of the space should be exhaustively sampled? And which part can be less examined? According to our results, it looks reasonable to sample along a possible binding pathway.

Although experimental data support the hypothesis that the portal (C2) is the entrance into the binding cavity of FABP4, the direct evidence has been lacking for a long time. The possibility that there exists another entrance cannot be excluded. A binding pathway through an aperture on the other side of the cavity, which corresponds to C4, has been found.[48] In our simulations, an orifice may appear in different locations (C3, C4, and C5) on the FABP4 surface, whereas the ligand does not penetrate into the protein through these sites and the only entrance observable is the portal. Due to limitation of the simulation length, we cannot completely rule out the possibility that the ligand may enter the binding cavity of FABP4 through another opening. However, the FES and the flux network clearly indicate that the pathway through the portal is energetically and kinetically favorable. From the present data we can infer that, even if the ligand is pulled out through another opening (for example, C3, C4, or C5) in the SMD simulation, the portal (C2) will be recognized to be the entrance into the binding cavity in the following MD simulations. In this sense our method can be employed to search for the preferred entry when the prior knowledge is not available.

In our simulations, except the internal cavity, a binding pocket on the protein surface is identified to be an allosteric site that may be applied in discovery of novel FABP4 inhibitors. With this approach the mechanism underlying ligand binding with different FABPs can be investigated. It will assist the development of specific inhibitors for an individual FABP, considering that the FABP family shares a conserved tertiary structure despite

diverse sequence similarity (22–73%). In FABP4 studies, Phe57 is regarded as the gating residue, which controls access of the ligand and water molecules to the interior cavity. In fact, aromatic residues work as a conformational gate in many proteins.[7] Therefore, our work is helpful in understanding the general mechanism of protein–ligand binding with a wing-type gate.

Method

Modeling of the simulation system

The FABP4–TGZ complex was obtained from PDB entry 2QM9.[45] The open conformation of FABP4 was obtained from PDB entry 3HK1.[49] The structure of TGZ was extracted and geometry optimized using Jaguar v7.9[50] with the B3LYP functional and the 6-31G* basis set. An electrostatic potential (ESP) for the ligand was generated by Jaguar. The atomic RESP charges were then determined by fitting with the RESP procedure implemented in Antechamber.[51]

The protein of FABP4 was modeled using the Amber ff03 force field,[52] and the ligand of TGZ was modeled using the general Amber force field (GAFF).[53] The protonation states of ionizable residues were determined by PROPKA implemented in Schrödinger Suite 2012 with pH of 7.0. The crystal structure was explicitly solvated in a rectangular box of $74 \times 71 \times 78$ Å^3 with 10183 TIP3P[54] water molecules. Chloride ions were added to neutralize uncompensated charges. Further salt (NaCl) was added to represent 0.15 M ionic concentration, mimicking the physiological environment. After the whole system was set up, a series of energy minimizations and equilibrations were performed. First, the water molecules, hydrogen atoms and salt ions were subjected to 3000 steps of steepest descent minimization followed by 12 000 steps of conjugate gradient minimization while other heavy atoms were constrained with a harmonic force of 2.0 kcal mol^{-1} Å^{-2}. Next, the whole system was energy minimized with 20 000 steps of the L-BFGS algorithm without any harmonic restraint. Then, coupled to a Langevin thermostat, the system was heated from 10 K up to 300 K by increments of 100 K in 20 ps and continued to run for 40 ps at 300 K at constant volume. Finally, the system was equilibrated for 200 ps in the *NPT* ensemble with the Langevin thermostat and isotropic position scaling, at 300 K and 1 bar.

The SMD and MD simulations were performed using Amber 11.[55] The equations of motion were solved with the leapfrog integration algorithm with a time step of 2 fs. The lengths of all bonds involving hydrogen atoms were kept constrained using the SHAKE algorithm.[56] The particle mesh Ewald (PME) method was applied for treating long-range electrostatic interactions. Periodic boundary conditions were used in all simulations.[57] A random seed was generated based on the current date and time to assign initial velocities.

Steered molecular dynamics (SMD)

The SMD simulation, which was extensively described,[31] was carried out using the SANDER program and the NCSU package implemented in Amber 11.[55] The reaction coordinate was defined as the linear combination of distances (LCOD) between heavy atoms of TGZ and Cys117. Cys117 was frozen with a restrain weight of 5.0 kcal mol^{-1} Å^{-2} in case it would rotate and point toward the protein surface in the simulation. With a harmonic constant of 2.0 kcal mol^{-1} Å^{-2} throughout the run, TGZ was pulled from the crystal position to the destination where the distance between TGZ and Cys117 was 45 Å. Thus the steering path was constructed. With the equilibrated system as the starting point, the SMD simulation lasted for 1 ns in the *NPT* ensemble with the Langevin thermostat and isotropic position scaling, at 300 K and 1 bar. The two hundred seeding structures were sampled at a time interval of 5 ps.

Molecular dynamics (MD)

The MD simulations were carried out using the parallel CUDA version of PMEMD on 2 GPUs.[58] For the dataset MD200, two hundred snapshot structures with water and ion molecules were extracted from the SMD trajectory. The water molecules and salt ions were subjected to 3000 steps of steepest descent minimization followed by 12 000 steps of conjugate gradient minimization while FABP4 and TGZ were frozen with a restrain weight of 10.0 kcal mol^{-1} Å^{-2} so that the relative position between them could not be changed. For each seeding structure, a 200 ns simulation was performed in the *NPT* ensemble using the Langevin thermostat and isotropic position scaling, at 300 K and 1 bar. For the dataset MD240, two hundred and forty snapshot structures with water and ion molecules were extracted from the MD trajectories of MD200. For each structure, a 400 ns simulation was performed in the *NPT* ensemble using the Langevin thermostat and isotropic position scaling, at 300 K and 1 bar. The trajectories were sampled at a time interval of 10 ps. Each snapshot structure was superimposed onto the crystal structure using the backbone heavy atoms of FABP4, and then the ligand RMSD with respect to the crystal structure and the coordinates of OAQ were obtained.

Bayesian clustering

Bayesian clustering, an unsupervised clustering method based on the Bayesian model, was performed using AutoClass C v3.3.6[59] in the three-dimensional space of OAQ's coordinates. The points extracted from MD simulations were classified. The clustering process started with a serial of seed numbers which gave an initial guess of the number of clusters. Then a random classification was generated and refined until a local maximum was found. This process was repeated 1000 times and the top two classifications with the highest probability were the output.

Hierarchical clustering

The hierarchical clustering was performed using the Ward algorithm implemented in R v3.0.[60] The dissimilarity between the classes obtained from Bayesian clustering was measured by the Euclidean distance between class means. The optimal number of clusters was determined by variation analysis of the height that measures the closeness of clusters. A smaller value of height, suggesting that the clusters are tightly grouped, is better. For both datasets (MD200 and MD240), six clusters were chosen because the height value decreases slowly with the number of clusters over six. The clustering results were checked by the distribution of the classes in the space to make sure that they are reasonable.

k-Means clustering

The k-means clustering was performed in the three-dimensional space of the OAQ's coordinates using the Hartigan–Wong algorithm implemented in R v3.0. Each of the six macrostates (C1–C6) and six microstates in C7 was clustered into twenty classes. From each class, the centroid point was chosen as the representative structure for following MD simulations.

2D free energy surface (FES)

The calculation of free energy is given by $W(r) = W_0 - k_B T \ln P(r)$, where W_0 is the depth of the FES, P is the probability distribution, k_B is the Boltzmann's constant, and T is the temperature. The coordinates of the atom OAQ sampled from MD simulations were projected onto the two-dimensional plane (xy, xz, and yz). The two-dimensional distribution function $P(r)$ was computed by assigning the data into 75×76 (MD200) and 75×79 (MD240) bins on the xy plane, 75×75 (MD200) and 75×80 (MD240) bins on the xz plane, and 76×75 (MD200) and 79×80 (MD240) bins on the yz plane with a bin area of 1.4×1.4 Å^2. As discussed previously,[8] the bin size had little effect on the FES.

Markov state model (MSM)

The MSM and TPT calculation was carried out using the software Emma v1.4.1.[61,62] The transition probability matrix was constructed with a sliding window of the lag time from 2 to 100 ns on each MD trajectory. The seven-state MSM was checked to be Markovian by analyzing the behaviors of the implied timescale. The implied timescale was calculated by $\tau_i = -\tau / \ln \lambda_i$, where λ_i is the eigenvalue of the transition matrix with the lag time of τ. With the lag time of 80 ns, the model was determined to be Markovian (ESI,† Fig. S5). Therefore, the transition probability matrix with the lag time of 80 ns was used in further computation of TPT and MFPT. The total flux from the unbound state C7 to the bound state C1 was decomposed into pathways between pairs of the seven states. The percentage contribution of each pathway to the total flux was computed.

The MFPT is defined as the mean time (f_i) needed to reach the final state for the first time from another state. It is obtained from the transition probability matrix by solving the following equation suggested previously.[63]

$$\begin{bmatrix} T_{1,1}-1 & \cdots & & & T_{1,m} \\ \vdots & T_{2,2}-1 & & & T_{2,m} \\ . & & \ddots & & \vdots \\ T_{m-1,1} & \cdots & & T_{m-1,m-1}-1 & T_{m-1,m} \\ 0 & 0 & 0 & & 1 \end{bmatrix} \times \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{m-1} \\ f_m \end{bmatrix} = \begin{bmatrix} -\tau \\ -\tau \\ \vdots \\ -\tau \\ 0 \end{bmatrix}$$

The statistical error in the MSM calculations was estimated by the standard deviation (SD) computed from three subsets of 180 MD trajectories taken from MD240 using the bootstrapping method. Each subset was constructed with 15 randomly chosen trajectories in each of the 12 important regions aforementioned.

The transition matrix and the fluxes between C1 and C7 were computed for each subset. The mean and SD are listed in ESI,† Tables S2–S6.

Calculation of binding affinity

With the seven-state MSM model, the binding affinity is calculated by $\Delta G = RT \ln(k_{off}/k_{on})$. The dissociation constant $k_{off} = 1/\text{MFPT}_{off}$ is directly achieved from MFPT. The association constant $k_{on} = 1/(\text{MFPT}_{on} \times C_{comp})$, where C_{comp} is the ligand concentration, depends on the ligand concentration and MFPT.

With the FES, the binding affinity is given by $\Delta G = -W_{3D} - k_B T \ln(V_b/V_0)$, where W_{3D} is the depth of the 3D FES, $V_b = \int_b \exp(-\beta W(r)) dr$ is the bound volume computed as the integral of the 3D FES, and $V_0 = 1661$ Å^3 is the volume of the standard state. The coordinates of the atom OAQ were assigned into $28 \times 28 \times 28$ (MD200) and $28 \times 30 \times 30$ (MD240) bins with a bin volume of $3.8 \times 3.8 \times 3.8$ Å^3. Then the three-dimensional FES was constructed and W_{3D} was estimated in the same way with the 2D FES. The bin size has little effect on the calculation of W_{3D}. When the bin volume varies from $3.0 \times 3.0 \times 3.0$ Å^3 to $5.0 \times 5.0 \times 5.0$ Å^3, the energy difference is less than 0.5 kcal mol^{-1}. For the dataset MD240, the similar bootstrapping method is applied to assess the error in the calculation of binding affinity (ESI,† Table S8).

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

References

1 M. Furuhashi and G. S. Hotamisligil, Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets, *Nat. Rev. Drug Discovery*, 2008, **7**, 489–503.

2 M. Furuhashi, G. Tuncman, C. Z. Gorgun, L. Makowski, G. Atsumi, E. Vaillancourt, K. Kono, V. R. Babaev, S. Fazio, M. F. Linton, R. Sulsky, J. A. Robl, R. A. Parker and G. S. Hotamisligil, Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2, *Nature*, 2007, **447**, 959–965.

3 K. M. Nieman, H. A. Kenny, C. V. Penicka, A. Ladanyi, R. Buell-Gutbrod, M. R. Zillhardt, I. L. Romero, M. S. Carey, G. B. Mills, G. S. Hotamisligil, S. D. Yamada, M. E. Peter, K. Gwin and E. Lengyel, Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth, *Nat. Med.*, 2011, **17**, 1498–1503.

4 R. T. Ogata, Kinetics of Fatty Acid Interactions with Fatty Acid Binding Proteins from Adipocyte, Heart, and Intestine, *J. Biol. Chem.*, 1996, **271**, 11291–11300.

5 J. Ory, C. D. Kane, M. A. Simpson, L. J. Banaszak and D. A. Bernlohr, Biochemical and crystallographic analyses of a portal mutant of the adipocyte lipid-binding protein, *J. Biol. Chem.*, 1997, **272**, 9793–9801.

6 A. E. Jenkins, J. A. Hockenberry, T. Nguyen and D. A. Bernlohr, Testing of the portal hypothesis: Analysis of a V32G, F57G, K58G mutant of the fatty acid binding protein of the murine adipocyte, *Biochemistry*, 2002, **41**, 2022–2027.

7 A. Gora, J. Brezovsky and J. Damborsky, Gates of Enzymes, *Chem. Rev.*, 2013, **113**, 5871–5923.

8 Y. Li, X. Li and Z. Dong, Concerted Dynamic Motions of an FABP4 Model and Its Ligands Revealed by Microsecond Molecular Dynamics Simulations, *Biochemistry*, 2014, **53**, 6409–6417.

9 D. A. Case and M. Karplus, Dynamics of Ligand-Binding to Heme-Proteins, *J. Mol. Biol.*, 1979, **132**, 343–368.

10 J. A. McCammon and S. H. Northrup, Gated Binding of Ligands to Proteins, *Nature*, 1981, **293**, 316–317.

11 H. X. Zhou, S. T. Wlodek and J. A. McCammon, Conformation gating as a mechanism for enzyme specificity, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 9280–9283.

12 J. A. McCammon, Gated Diffusion-controlled Reactions, *BMC Biophys.*, 2011, **4**, 4.

13 R. Elber and M. Karplus, Enhanced Sampling in Molecular-Dynamics - Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon-Monoxide Diffusion through Myoglobin, *J. Am. Chem. Soc.*, 1990, **112**, 9161–9175.

14 M. A. Mart, M. C. G. Lebrero, A. E. Roitberg and D. A. Estrin, Bond or cage effect: How nitrophorins transport and release nitric oxide, *J. Am. Chem. Soc.*, 2008, **130**, 1611–1618.

15 J. Z. Ruscio, D. Kumar, M. Shukla, M. G. Prisant, T. M. Murali and A. V. Onufriev, Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9204–9209.

16 R. Baron, C. Riley, P. Chenprakhon, K. Thotsaporn, R. T. Winter, A. Alfieri, F. Forneris, W. J. H. van Berkel, P. Chaiyen, M. W. Fraaije, A. Mattevi and J. A. McCammon, Multiple pathways guide oxygen diffusion into flavoenzyme active sites, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10603–10608.

17 L. Maragliano, G. Cottone, G. Ciccotti and E. Vanden-Eijnden, Mapping the Network of Pathways of CO Diffusion in Myoglobin, *J. Am. Chem. Soc.*, 2010, **132**, 1010–1017.

18 M. S. Shadrina, A. M. English and G. H. Peslherbe, Effective Simulations of Gas Diffusion Through Kinetically Accessible Tunnels in Multisubunit Proteins: O-2 Pathways and Escape Routes in T-state Deoxyhemoglobin, *J. Am. Chem. Soc.*, 2012, **134**, 11177–11184.

19 P. H. Wang and J. Blumberger, Mechanistic insight into the blocking of CO diffusion in NiFe -hydrogenase mutants through multiscale simulation, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 6399–6404.

20 P. H. Wang, M. Bruschi, L. De Gioia and J. Blumberger, Uncovering a Dynamically Formed Substrate Access Tunnel in Carbon Monoxide Dehydrogenase/Acetyl-CoA Synthase, *J. Am. Chem. Soc.*, 2013, **135**, 9493–9502.

21 F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff and P. A. Anfinrud, Watching a Protein as it Functions with 150-ps Time-Resolved X-ray Crystallography, *Science*, 2003, **300**, 1944–1947.

22 D. Dantsker, U. Samuni, J. M. Friedman and N. Agmon, A hierarchy of functionally important relaxations within myoglobin based on solvent effects, mutations and kinetic model, *Biochim. Biophys. Acta, Proteins Proteomics*, 2005, **1749**, 234–251.

23 R. Elber, Ligand diffusion in globins: simulations *versus* experiment, *Curr. Opin. Struct. Biol.*, 2010, **20**, 162–167.

24 I. Buch, T. Giorgino and G. De Fabritiis, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10184–10189.

25 R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu and D. E. Shaw, Pathway and mechanism of drug binding to G-protein-coupled receptors, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13118–13123.

26 D. Huang and A. Caflisch, The Free Energy Landscape of Small Molecule Unbinding, *PLoS Comput. Biol.*, 2011, **7**, e1002002.

27 Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger and D. E. Shaw, How Does a Drug Molecule Find Its Target Binding Site?, *J. Am. Chem. Soc.*, 2011, **133**, 9181–9183.

28 K. M. ElSawy, R. Twarock, D. P. Lane, C. S. Verma and L. S. D. Caves, Characterization of the Ligand Receptor Encounter Complex and Its Potential for in Silico Kinetics-Based Drug Development, *J. Chem. Theory Comput.*, 2012, **8**, 314–321.

29 D. Nerukh, N. Okimoto, A. Suenaga and M. Taiji, Ligand Diffusion on Protein Surface Observed in Molecular Dynamics Simulation, *J. Phys. Chem. Lett.*, 2012, **3**, 3476–3479.

30 Y. Li, X. Li, W. Ma and Z. Dong, Conformational Transition Pathways of Epidermal Growth Factor Receptor Kinase Domain from Multiple Molecular Dynamics Simulations and Bayesian Clustering, *J. Chem. Theory Comput.*, 2014, **10**, 3503–3511.

31 S. Park, F. Khalili-Araghi, E. Tajkhorshid and K. Schulten, Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality, *J. Chem. Phys.*, 2003, **119**, 3559–3566.

32 A. Shinobu and N. Agmon, The Hole in the Barrel: Water Exchange at the GFP Chromophore, *J. Phys. Chem. B*, 2015, **119**, 3464–3478.

33 Y. Li, Bayesian model based clustering analysis: Application to a molecular dynamics trajectory of the HIV-1 integrase catalytic core, *J. Chem. Inf. Model.*, 2006, **46**, 1742–1750.

34 F. Pietrucci, F. Marinelli, P. Carloni and A. Laio, Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations, *J. Am. Chem. Soc.*, 2009, **131**, 11811–11818.

35 N.-j. Deng, W. Zheng, E. Gallicchio and R. M. Levy, Insights into the Dynamics of HIV-1 Protease: A Kinetic Network Model Constructed from Atomistic Simulations, *J. Am. Chem. Soc.*, 2011, **133**, 9387–9394.

36 M. Held, P. Metzner, J.-H. Prinz and F. Noe, Mechanisms of Protein-Ligand Association and Its Modulation by Protein Mutations, *Biophys. J.*, 2011, **100**, 701–710.

37 S. K. Sadiq, F. Noe and G. De Fabritiis, Kinetic characterization of the critical step in HIV-1 protease maturation, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20449–20454.

38 C. R. Schwantes and V. S. Pande, Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9, *J. Chem. Theory Comput.*, 2013, **9**, 2000–2009.

39 K. J. Kohlhoff, D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman and V. S. Pande, Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways, *Nat. Chem.*, 2014, **6**, 15–21.

40 G. Adam and M. Delbruck, Reduction of dimensionality in biological diffusion processes, in *Structural chemistry and molecular biology*, ed. A. Rich, N. Davidson and W. H. Freeman and Co., San Francisco, 1968, pp. 198–215.

41 A. J. Smith, B. R. Thompson, M. A. Sanders and D. A. Bernlohr, Interaction of the adipocyte fatty acid-binding protein with the hormone-sensitive lipase - Regulation by fatty acids and phosphorylation, *J. Biol. Chem.*, 2007, **282**, 32424–32432.

42 A. J. Smith, M. A. Sanders, B. E. Juhlmann, A. V. Hertzel and D. A. Bernlohr, Mapping of the Hormone-sensitive Lipase Binding Site on the Adipocyte Fatty Acid-binding Protein (AFABP) Identification of the Charge Quartet on the AFABP/AP2 Helix-turn-helix Domain, *J. Biol. Chem.*, 2008, **283**, 33536–33543.

43 N. Agmon and J. J. Hopfield, CO binding to heme proteins: A model for barrier height distributions and slow conformational changes, *J. Chem. Phys.*, 1983, **79**, 2042–2053.

44 N. Agmon, Conformational Cycle of a Single Working Enzyme, *J. Phys. Chem. B*, 2000, **104**, 7830–7834.

45 R. E. Gillilan, S. D. Ayers and N. Noy, Structural basis for activation of fatty acid-binding protein 4, *J. Mol. Biol.*, 2007, **372**, 1246–1260.

46 E. Gallicchio and R. M. Levy, Advances in all atom sampling methods for modeling protein-ligand binding affinities, *Curr. Opin. Struct. Biol.*, 2011, **21**, 161–166.

47 S. Doudou, N. A. Burton and R. H. Henchman, Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force, *J. Chem. Theory Comput.*, 2009, **5**, 909–918.

48 R. Friedman, E. Nachliel and M. Gutman, Molecular dynamics simulations of the adipocyte lipid binding protein reveal a novel entry site for the ligand, *Biochemistry*, 2005, **44**, 4275–4283.

49 A. V. Hertzel, K. Hellberg, J. M. Reynolds, A. C. Kruse, B. E. Juhlmann, A. J. Smith, M. A. Sanders, D. H. Ohlendorf, J. Suttles and D. A. Bernlohr, Identification and Characterization of a Small Molecule Inhibitor of Fatty Acid Binding Proteins, *J. Med. Chem.*, 2009, **52**, 6024–6031.

50 *Jaguar, version 7.9*, Schrödinger, LLC, New York, NY, 2012.

51 Antechamber. AmberTools Users' Manual, version 1.5.

52 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

53 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

54 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

55 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, The Amber biomolecular simulation programs, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

56 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical-integration of cartesian equations of motion of a system with constraints – molecular-dynamics of *n*-alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.

57 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh ewald method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

58 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.

59 D. Cook, J. Potts, W. Taylor, AutoClass C, http://ti.arc.nasa.gov/tech/rse/synthesis-projects-applications/autoclass/autoclass-c/.

60 R. The R Project for Statistical Computing, http://www.r-project.org/.

61 F. Noe and S. Fischer, Transition networks for modeling the kinetics of conformational change in macromolecules, *Curr. Opin. Struct. Biol.*, 2008, **18**, 154–162.

62 M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schuette and F. Noe, EMMA: A Software Package for Markov Model Building and Analysis, *J. Chem. Theory Comput.*, 2012, **8**, 2223–2238.

63 N. Singhal and V. S. Pande, Error analysis and efficient sampling in Markovian state models for molecular dynamics, *J. Chem. Phys.*, 2005, **123**, 204909.

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 32257–32267 | 32267