Mechanisms of protein-ligand association and its modulation by protein mutations

Martin Held[1,4], Philipp Metzner[2,3], Jan-Hendrik Prinz[1,3], and Frank Noé[1,3]

[1]Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[2]University of Lugano, Via Giuseppe Buffi 13, 6900 Lugano, Switzerland
[3]DFG Research Center MATHEON
[4]IMPRS-CBSC

January 8, 2011

Abstract

Protein-ligand interactions are essential for nearly all biological processes, and yet the biophysical mechanism that enables potential binding partners to associate before specific binding occurs remains poorly understood. Fundamental questions include which factors influence the formation of protein-ligand encounter complexes, and whether designated association pathways exist. In this article we introduce a computational approach to systematically analyze the complete ensemble of association pathways and to thus investigate these questions. This approach is employed here to study the binding of a phosphate ion to the *Escherichia coli* Phosphate Binding Protein. Various mutants of the protein are considered and their effects on binding free energy profiles, association rates and association pathway distributions are quantified. The results reveal the existence of two *anion attractors*, i.e. regions that initially attract negatively charged particles and allow them to be efficiently screened for phosphate which is specifically bound subsequently. Point mutations that affect the charge on these attractors modulate their attraction strength and speed up association to a factor of 10 of the diffusion limit and thus change the association pathways of the phosphate ligand. It is demonstrated that a phosphate that pre-binds to such an attractor neutralizes its attraction effect to the environment, making the simultaneous association of a second phosphate ion unlikely. Our study suggests ways how structural properties can be used to tune molecular association kinetics so as to optimize the efficiency of binding, and highlights the importance of kinetic properties.

Keywords: protein-ligand binding, association pathways, Phosphate Binding Protein, Brownian dynamics, Transition Path Theory

1 Introduction

The ability of proteins to bind ligands, including ions, substrates, co-factors and other proteins, is essential to all life processes. For instance, protein-ligand interaction mediates uptake and storage of cargo such as oxygen uptake in Hemoglobin, molecular recognition leading to information transfer such as in sensing of neurotransmitters or growth hormones, and buildup of biological structures such as in RNA-Ribosome interactions (1, 2). While the majority of the biochemical and pharmaceutical work investigated protein-ligand interactions in terms of equilibrium binding affinities, it is becoming increasingly evident that the effectiveness of such interactions crucially depends on dynamical and kinetic properties (3). The dynamical properties of binding are inherently linked to structural aspects such as size, concentration and spatial distribution of the binding partners as well as their detailed atomic structures and changes therein.

Structure-dynamics relationships for binding processes have been studied a lot at binding site contact distance, both on relevant energetics such as detailed electrostatic complementarity of the binding surfaces and hydrophobic burial, as well as on the structural binding mechanisms such as induced fit versus conformational selection (4, 5). In contrast, fundamental properties of the spatiotemporal mechanism of how this first contact of the protein-ligand binding process is established are still elusive. For example, does binding occur via a single dominant pathway, via

multiple separated pathways, or via a funnel-shaped ensemble of pathways? Is it directed to the binding site or do metastable states exist which trap the binding partners in nonfunctional states? Can diffusion-limited binding be sped up by rapid binding to the surface and subsequent surface search?

From a theoretical point of view, the protein-ligand association process can be thought of as a diffusion in a high dimensional energy landscape that usually has an energetically favorable minimum at the configuration of the protein-ligand complex. In situations in which the interaction process takes place in dilute media, this energy landscape is flat at large protein-ligand distances, resulting in a purely diffusive motion of the molecules. When the interaction partners approach each other, electrostatic forces become relevant and for favorably interacting molecules, the energy landscape funnels down towards the complex formation configuration (6). Such a binding funnel may also possess complex features as local minimal or parallel pathways. All mechanistic questions can be answered when the binding funnel and the dynamics governing the motion upon it are understood. Protein-ligand binding has thus many similarities with protein folding and principles or methods worked out in the protein folding field are also likely to be useful here.

In the past decades, the field of molecular simulations has been increasingly successful assigning structural and mechanistical information to experimental observations (7). A widely used computational approach to simulate protein-ligand association dynamics are Brownian dynamics (BD) and Langevin dynamics (LD) simulations (8) of the diffusional motion of internally rigid protein models in implicit solvent. The BD approach has been proven useful to predict bi-molecular association rates (9–12) in situations where binding is diffusion limited, as well as to provide detailed insights into how protein-ligand encounter complexes are formed (13). However, a systematic analysis of the obtained simulation data is often difficult. In this work we present a simulation and analysis approach that directly reveals the ensemble of pathways of a ligand to the binding pocket, thus allowing mechanistic questions to be answered. The approach allows to identify the presence of metastable states in the binding procedure and to study how binding mechanism and rates are altered by mutations in the protein.

Two alternative approaches to simulate and analyze dynamics exist: Most commonly, one uses the *direct simulation* approach in which long trajectory realizations of the dynamical equations (e.g. BD) are generated, and then analyzed. This approach has the advantage that it allows complex geometries with many degrees of freedom to be simulated, such as large heterogeneous protein mixtures (14). A disadvantage is that quantities computed from generated trajectories, e.g., association rates come with statistical uncertainty, or may be systematically biased when some rare events have not been sampled at all. Moreover, trajectory data are often tedious to analyze, involving the search for "interesting" observables which involves human subjectivity. Alternatively, one can describe the *ensemble dynamics* of the system, where the transition probabilities or rates between configurational substates of the system are characterized. This approach has been successfully used in modeling the conformational dynamics of proteins with Markov models (15–19), where the interstate transition probabilities are estimated from many short simulations that are initialized from different substates. In diffusion processes, such as BD and LD, the ensemble dynamics can be expressed directly via the Fokker-Planck equation. Based on this formulation, sampling of individual trajectories can be avoided and the sampling error can be made zero. However, the downside of this approach is that for solving the Fokker-Planck equation the configuration space must be discretized. When using a rectangular lattice, this is currently only feasible for three-dimensional spaces. Nevertheless, with a three-dimensional space one can already address the biophysically interesting process of ion binding to proteins (20). An extension to higher-dimensional problems such as protein-protein binding with internal dynamics is feasible by extending the approach to meshless discretization approaches (17, 21, 22).

In this paper we show how the *ensemble dynamics* approach permits a straightforward and objective analysis of the protein-ligand association pathways by using the mathematical framework of Transition Path Theory (TPT) (23, 24) which provides a complete and quantitative description of association pathways that lead from a freely diffusing ligand towards a protein-ligand complex in a given molecular model. Here, we apply this approach to systematically study the binding of inorganic phosphate (P_i) to the Phosphate Binding Protein (PBP) of *Escherichia coli* (25–27).

This protein plays an important role in the phosphate supply of bacterial cells and is expressed in situations when the intracellular concentration of P_i is low. After it is transported to the bacterial periplasm, it scavenges for free P_i to subsequently pass it on to a membrane protein which transports the phosphate into the cytoplasm. While previous work on the binding of P_i to PBP was mainly concerned with investigating the binding kinetics by experimental means (28, 29) or direct simulation (30), to our knowledge this work for the first time provides a systematic description of the P_i binding pathway ensemble. Various mutations are studied and it is shown how they modulate the phosphate binding rates and pathways. It is also shown how PBP becomes saturated to a second binding attempt once a P_i has been bound.

The obtained findings highlight the importance of a positively charged patch of the PBP for the attraction of negatively charged ions. Our results suggest that this pre-binding site may be a general mechanism for efficiently organizing specific ion binding via a two-step mechanism that first selects by polarity and then by ion type.

2 Theory

2.1 Dynamical Model

Without loss of generality, the protein-ligand association process can be divided into two phases that are dominated by different forces (31) (see Figure 1). The association phase I is largely governed by electrostatic forces and thermal motion of solvent molecules which lead to a diffusive approach of the solutes studied, and does not depend on intramolecular flexibility. In the binding phase II the protein-ligand complex is formed, which involves more complex short range forces, intramolecular flexibility and the structural role of solvent molecules. This separation into two phases suggests two different computational models to describe them. The second phase requires a more detailed approach such as all-atom molecular dynamics simulation with full structural resolution and flexibility. Here, we restrict ourselves on the association phase I where the motion of the ligand in the protein-ligand potential is described by rigid body Brownian (or Smoluchowski) dynamics in implicit solvent:

$$d\mathbf{x}(t) = -\frac{D}{k_B T}\,\boldsymbol{\nabla} V(\mathbf{x})\,dt + \sqrt{2D}d\mathbf{W}_t, \tag{1}$$

where $\mathbf{x}(t)$ is the position of the ligand at time $t \geq 0$, D is the joint translational diffusion constant of PBP and P_i, T the absolute temperature, k_B the Boltzmann constant, $V(\mathbf{x})$ the potential energy of the ligand at position $\mathbf{x}(t)$ and $\mathbf{W}_t$ is multivariate Wiener process, i.e. white noise with independent, normally distributed, increments. We assume isotropic diffusion for both the protein and the ligand, hence diffusion can be described by a scalar constant. The error introduced by neglecting hydrodynamic interactions between interaction partners is unlikely to affect the main findings of this paper. However, in subsequent studies hydrodynamic interaction could be included following recent work of Geyer et. al (32). The change in particle position $d\mathbf{x}(t)$ in a time interval dt is thus the result of the force from the potential, $-\nabla V(\mathbf{x})$, and a random displacement which implicitly models the collisions with solvent molecules. It is important to note that the solution $x(t)$ of the stochastic differential equation (Eq. 1) is a random sequence. Hence, for a given initial position $\mathbf{x}(0) = \mathbf{x}_o$, each realization of $\mathbf{x}(t)$ describes a possible ligand trajectory.

2.2 Interaction Potential

In order to compute the interaction potential between PBP and phosphate ion solely electrostatic forces are considered as they are the most important contributors during the association phase. An explicit modeling of Van der Waals forces can be omitted as the interaction partners can be thought of being immersed in dense media (water) and therefore interact equally with all surrounding atoms.

Furthermore, the structure of the diffusing ligand P_i is approximated by a point charge of $-2e$ to represent the HPO_2^{-2} form of phosphate. This allows the energy of PBP-P_i configurations to

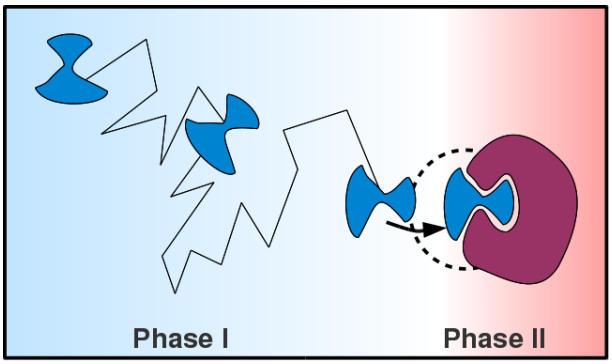

Figure 1: Phases of protein-ligand association. Phase I is largely characterized by diffusional association and ends upon encounter complex formation. Phase II involves actual binding of the ligand and might involve structural rearrangements in both interaction partners.

be calculated by multiplying the electrostatic potential induced by the protein with the phosphate charge ($-2e$) at the respective positions. The protein potential $V(\mathbf{r})$ is computed using Poisson-Boltzmann theory (33), in which the solvent is modeled as a continuum with a specific dielectric constant. The Poisson-Boltzmann equation is given by:

$$\boldsymbol{\nabla}[\epsilon(\mathbf{x})\boldsymbol{\nabla}V(\mathbf{x})] = -4\pi\rho(\mathbf{x}) - 4\pi\sum_i c_i^\infty z_i q \exp\left(\frac{-z_i q V(\mathbf{x})}{k_B T}\right)\lambda(\mathbf{x}) \tag{2}$$

where $\epsilon(\mathbf{x})$ is the dielectric constant at position $\mathbf{x}$, ρ indicates the charge density of the protein (given by assigning partial atom charges), c_i^∞ denotes the concentration of ion species i at an infinite distance from the molecule, z_i is its partial charge, q is the elementary charge, k_B the Boltzmann constant, T is the temperature and $\lambda(\mathbf{x})$ indicates the ion accessibility.

For the calculation of association rates to be correct, the volume considered around the protein has to be large enough such that the gradient of the potential approaches zero at its outer boundaries. At the same time it is crucial for a correct calculation that the potential close to the protein surface is well described. To comply with the large volume and high resolution requirements, we use the manual focusing mechanism (mg-manual) provided by APBS, solving the PB equation on differently sized grids ranging from $33 \times 33 \times 33$ with isotropic spacing of $d = 16$ Å to $193 \times 193 \times 193$ with isotropic spacing of $d = 0.35$ Å. The respective coarser solutions was used as an outer boundary condition for the finer one.

2.3 Transition Path Theory

While individual realizations of the stochastic dynamics (Eq. 1) are random, we are interested in the deterministic expectation values of this random process, such as transition rates, fluxes and pathway probabilities. In order to obtain these quantities, we apply Transition Path Theory (TPT) (23, 34, 35) to the Markov jump process that results from discretizing the Fokker-Planck equation that is associated with Eq. 1.

The main concepts of TPT are briefly restated subsequently. Given an ergodic stochastic process such as Brownian dynamics in a potential, Langevin dynamics or a Markov jump process, e.g., on a grid, TPT provides the statistical properties of the ensemble of reactive pathways between two disjoint subsets, "A" and "B", of the state space. For this, consider a hypothetical infinitely long trajectory. A trajectory fragment is called a reactive trajectory, if it leaves A and subsequently enters B. Particularly, trajectories that return to A before reaching B are not considered part of the reactive trajectory ensemble. (See Figure 2 for an illustration)

In order to calculate TPT quantities in practice the configuration space must be discretized. We consider that the configuration space is partitioned into small sets, here briefly called "states".

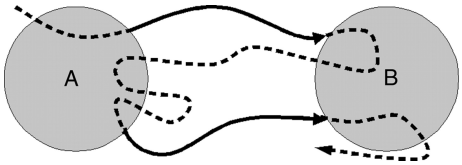

Figure 2: Reactive trajectory definition. The bold parts show reactive trajectory parts going from set A (unbound configurations) to set B (bound configurations)

In the scenario of protein-ligand binding, the set A is defined to comprise of configurations where the ligand can freely diffuse, here chosen to contain all states in space that are more than 250 Å away from the center of mass of the protein. The set B in turn is chosen to contain all states that correspond to bound or metastable pre-complex protein-ligand configurations.

The essential quantity needed to compute statistical properties of transition pathways between A and B is the forward committor, q_i^+, defined as the probability that the process when being at state i will reach the set B next, rather than returning to set A. In the context of protein-ligand association, q_i^+denotes the probability to associate to the binding site (at B) rather than to dissociate to set A. In the methods section it is explained how the forward commitor q_i^+ can be efficiently calculated for a given dynamical model. Furthermore, we need the backward committor probability q_i^-, which is the probability when being at state i that the process has been in set A previously rather than in set B, i.e., it has reached state i from the dissociated states A and not been bound before. For reversible stochastic processes, as in the present case, it is simply given by $q_i^- = 1 - q_i^+$. Let k_{ij} be the transition rate between states i and j, without taking into account the choice of A and B. In order to be able to infer information about the reactive parts of the trajectory, i.e., the parts that leave A and go to B, only the part of k_{ij} must be considered, which involves trajectories that come from the dissociated A set and will go on to the associated B set, i.e., $q_i^- k_{ij} q_j^+$. The reactive probability flux is hence given by

$$f_{ij} = \pi_i q_i^- k_{ij} q_j^+. \tag{3}$$

π_i denotes the Boltzmann weight of state i, i.e., the overall probability for the process to be in the volume element represented by state i. This definition still contains recrossing events of reactive trajectories. In order to only account for the net reactive probability flux from A to B, the reactive flux f_{ji} associated with recrossings of the reactive trajectory is subtracted from the forward flux f_{ij}, leading the following expression for the net reactive probability flux:

$$f_{ij}^+ = \max\{0, f_{ij} - f_{ji}\}. \tag{4}$$

It is important to note that the flux is conserved, i.e., the amount of flux leaving A equals the amount entering B and for all intermediate states i the influx equals the outflux. This property directly leads to an expression for the transition rate from A to B which is explained in the next section. Refer to Figure 3 for an illustration of TPT on a two-dimensional model of protein-ligand association.

2.4 Binding Rate Calculation

The expected number of $A \to B$ transitions per time unit is given by the total flux (36):

$$F_{AB} = \sum_{i \in A} \sum_{j \notin A} f_{ij}^+ = \sum_{i \in A} \sum_{j \notin A} \pi_i k_{ij} q_j^+. \tag{5}$$

This quantity includes the fact that the ligand needs to diffuse back to the A area until another transition to B is considered. Hence, in order to calculate the $A \to B$ transition rate, we need to take the probability into account, that the ligand is moving from A to B, i.e., it has been in A last:

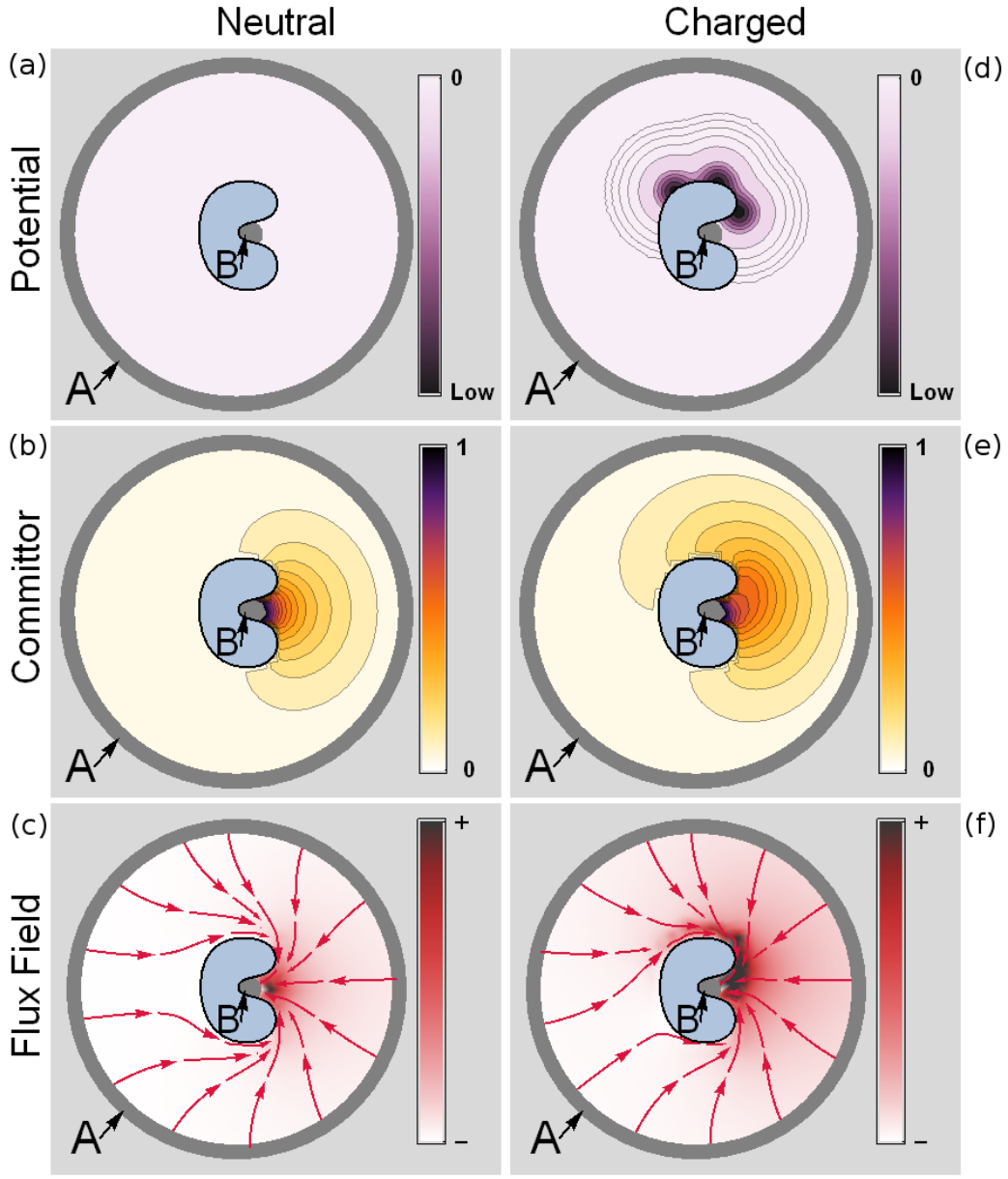

Figure 3: Illustration of TPT on a simple two-dimensional protein-ligand binding model. The dissociated state of the ligand A and the associated state B are shown. (a), (b), (c) show a situation in which no potential is present and the ligand can diffuse freely until it associates to the protein. (d), (e), (f) illustrate a situation where the protein has surface charges, generating energy minima that attract the ligand. (a) and (d) show the different potentials, (b), (e) show the forward committor q^+, revealing for areas on top of the charged protein a higher probability to reach the binding site than for the uncharged protein. (c) and (f) show the reactive flux density and integrated flux lines calculated from the flux field resulting from the fluxes f_{ij}^+. For the neutral protein it becomes apparent that the ligand diffuses freely and the binding pathways are only restricted by spatial constraints. In the charged scenario the ligand is strongly attracted by the top side of the protein, creating a high reactive flux density in that area that distorts the pathway ensemble accordingly.

$$\pi_A = \sum_{i \in S} \pi_i q_i^-, \tag{6}$$

where S is the set of all states. Therefore, the transition rate is given by (15):

$$k_{AB} = \frac{F_{AB}}{\pi_A}. \tag{7}$$

k_{AB} is the rate at which a ligand molecule binds starting from set A. In order to compute the bimolecular association rate of PBP and P_i, the rate at which ligand molecules arrive at the A sphere has to be taken into account. Based on the assumption that protein and ligand diffuse freely upon a distance r, i.e., in our scenario the ligand enters the A sphere, according to Erban et al. (37) the diffusion limited association constant k_{On} can be obtained by:

$$k_{\mathrm{On}} = 4\pi D \left(r - \sqrt{\frac{D}{k_{AB}}} \tanh\left(r \sqrt{\frac{k_{AB}}{D}} \right) \right), \tag{8}$$

where D is the diffusion constant, and r denotes the radius of the A sphere. Note that k_{On} is a concentration dependent rate (e.g. in $nm^3 s^{-1}$), while k_{AB} is the rate of a single molecule event (in s^{-1}).

3 Methods

3.1 Molecular model and simulation setup

The coordinates of the open form mutant T141D of the Phosphate Binding Protein from *Escherichia coli* (PDB (38) code 1OIB, Chain A) served as a template to create several *in silico* mutants of the protein. The mutagenesis tool of PyMOL (vers. 0.99rc6) was used to create mutants D56N, D137T, K43M, K43Q, R134Q, R135Q, R134Q/K167Q/K175Q (3 mut.), R134Q/K167Q/K175Q/D21N/D51N/D61N (6 mut.), T141D, chosen in agreement with previous work on PBP (30). The wild-type (wt) was modeled by replacing Asp141 with Thr141.

Energy minimization of the structures in a TIP3P water box was carried out by running 2000 steps of the steepest gradient algorithm using the Gromacs (version 4.5) program (39) employing the CHARMM (40) force-field. The protonation states of ionizable amino acids were determined by using the PROPKA (41) tool setting the pH to 7. The atomic partial charges were assigned using the PDB2PQR suite (42) using the CHARMM force-field as reference. The electrostatic potential of the resulting structures were calculated using the Adaptive Poisson-Boltzmann Solver (APBS) (43), using dielectric constants of $\epsilon_P = 4.0$ for the protein interior and $\epsilon_S = 78.0$ for the solvent. As joint diffusion constant $D = 8 \times 10^{-6} cm^2 s^{-1}$ (44) was used.

3.2 Space Discretization

In order to calculate the TPT quantities for the protein-ligand binding process, a finite volume space discretization is required which extends over a large volume while at the same time having a high resolution close to the protein surface. Therefore, we developed a simple adaptive discretization scheme based on the numerical gradient of the electrostatic potential. The procedure starts from a coarse cubic $33 \times 33 \times 33$ grid with an edge length of 528 Å and refines interior grid points based on a local error criterion. By using central finite differences the potential derivatives in each Euclidean direction are computed for each point, at the one hand using the present discretization as well as using a finer discretization where additional grids points have been added halfway between each pair of initial grid points. Whenever at a given refinement point the two derivatives differ by more than a specified threshold (here 0.01 $kT/\mathring{A}$), the refinement is accepted and another grid plane is added intersecting with this refinement point and perpendicular to the connection between the two coarse grid points. This procedure is iterated until no more planes are added.

Grid points that would lie inside the protein, defined by having a minimal distance to protein atoms of less than 3.2 Å, are not taken into account, and are dismissed from the final grid. The resulting grids had an average size of $173 \times 151 \times 177$ points (a total of 4.623.771 elements) with box lengths ranging from 16 Å for distant boxes to 0.5 Å in the vicinity of the protein.

3.3 Rate Matrix Computation

When considering Brownian dynamics (Eq. 1) the transition rates between volume elements of the regular grid defined in 3.2 can be computed using a discretization scheme introduced in (45). The resulting matrix $\mathbf{K}$ is a discrete model of the entire ensemble dynamics of the protein-ligand association process, and all subsequent analysis can be conducted based on this matrix. A matrix element k_{ij} specifies the number of transitions per time unit to a volume element j conditional on starting at element i, and is computed as follows:

$$k_{ij} = \begin{cases} \frac{D}{h_{i,j} d_{i \to j}} \exp\left(-\frac{1}{2 k_B T}(V_j - V_i)\right) & j \in \{N_i\} \\ -\sum_j k_{ij}, & j = i \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where N_i denotes the set of all volume elements that share a face with element i, D is the joint diffusion constant, V_i designates the potential at grid point i, $h_{i,j}$ denotes the distance between grid points i and j and $d_{i \to j}$ stands for the length of the ith volume cell into the direction of j.

3.4 A and B Definition and Committor Computation

After obtaining the space discretization of the volume around the protein, the A and B sets are assigned. For the set of free diffusing configurations of the phosphate ion (A set of states) all volume elements whose center is further than 250 Å away from the geometric center of the protein are chosen. Note that the choice of A is irrelevant as long as far enough away from the protein such that the electrostatic forces are zero in A. Defining A further away from this minimal distance will increase r but decrease k_{AB} resulting in the same concentration dependent binding rate as in Eq. 8. The set B of bound/pre-complex configurations is chosen to include all volume elements that are within a 3 Å radius of the geometric center of Thr10, Ser38 and Ser139, shown as yellow region in Figure 4. The choice of B will affect the pathways and association rates as it defines the bound state.

With the discrete rate matrix (Eq. 9) the forward committor can be computed by solving the constrained linear problem

$$\begin{aligned} \mathbf{K}\mathbf{q} &= 0 \\ \text{s.t.} \quad & q_i = 0 \, \forall i \in A \\ & q_i = 1 \, \forall i \in B \end{aligned} \tag{10}$$

where A and B are the sets of discrete states corresponding to dissociated and associated states, respectively. This problem is solved by reordering the states in the order (S, A, B) where $S = (A \cup B)^C$, yielding the following structure in $\mathbf{K}$ and $\mathbf{q}$:

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{SS} & \mathbf{K}_{SA} & \mathbf{K}_{SB} \\ \mathbf{K}_{AS} & \mathbf{K}_{AA} & \mathbf{K}_{AB} \\ \mathbf{K}_{BS} & \mathbf{K}_{BA} & \mathbf{K}_{BB} \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} \mathbf{q}_S \\ \mathbf{q}_A = \mathbf{0} \\ \mathbf{q}_B = \mathbf{1} \end{pmatrix}, \tag{11}$$

Which allows Eq. (10) to be rewritten as:

$$\mathbf{K}_{SS}\mathbf{q}_S = \mathbf{K}_{SB}. \tag{12}$$

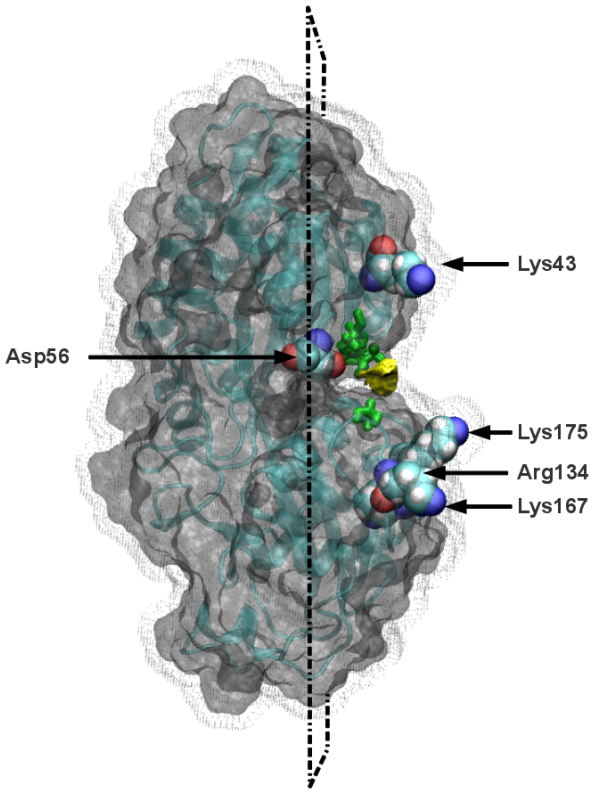

Figure 4: Transparent Connolly surface of the Phosphate Binding Protein (*Escherichia coli*) showing secondary structure elements. The yellow region depicts the B set. A subset of mutated amino acids are shown in Van der Waals representation. The dotted surface represents points acccessible by the phosphate ion. The indicated plane denotes the projection area used to visualize first hitting densities (Figs. 6 and 7).

which can easily be solved by standard numerical methods to obtain the unknown $\mathbf{q}_S$. In the present application the number of unknowns is in the order of 10^6. In order to solve this task we use the implementation of the iterative BiCGStab algorithm provided by the Java Matrix Tookit (46). A thorough discussion of efficient committor computations including error analysis can be found in (47).

3.5 Free Energy Profile of Ligand Association

As the forward committor is the probability to associate rather than dissociate, it measures the progress of the reaction and thus represents a "kinetic reaction coordinate" (48) with 0 representing the dissociated (A) and 1 representing the associated configurations (B). The free energy along this coordinate is given by

$$F(q) = -k_B T \ln(\rho(q)) + const. \tag{13}$$

$\rho(q)$ denotes the stationary density of the set of states having a committor value q and is calculated in our discrete model by

$$\rho(q) = \sum_{i,\, q_i \in [q-\frac{\Delta}{2}, q+\frac{\Delta}{2}]} \exp\left(\frac{-V(x_i)}{k_B T}\right). \tag{14}$$

using a sliding window with width $\Delta = 0.005$ over the range of $q \in [\frac{\Delta}{2}, 1-\frac{\Delta}{2}]$.

3.6 Binding flux field and visualization

For a visualization of phosphate association pathways a vector field of reactive fluxes was calculated. For this purpose, a total flux vector was assigned to each grid point i by vectorial summation of all outgoing fluxes f_{ij}^{+}. To visualize the resulting vector field, as in Figs. 6 and 7, the Mayavi2 program (49) was employed. Starting from a fixed number of points spherically distributed with distance 80 Å from the geometric center of the protein, the program follows the streamlines along the flux vectors, thus tracing out possible binding pathways. The streamlines are colored according to the local flux strength, i.e., norm of the total flux vectors. The lighter the coloring, the stronger the encountered flux.

In order to better visualize how the association pathways behave near the protein, we have calculated where they hit the protein surface for the first time. For this, a surface was defined in a distance of 10 Å around the phosphate accessible surface. At each surface element, the flux through the surface, quantified by the reactive TPT flux f_{ij}^{+} (Eq. 4), is calculated. For the sake of visualization the orthogonal projection of surface elements onto a two-dimensional plane which divides the surface into two halves is calculated. The plane is depicted in Figure 4. In the projection only surface elements on the half of the binding site are taken into account.

4 Results and Discussion

The results of the modeling and analysis of inorganic phosphate association to the phosphate binding protein and various *in silico* mutants are presented below. Selected mutants are summarized in Figure 6 while the results for remaining structures are shown in the supplementary information (SI).

4.1 Free Energy Profiles and Association Rates

The left column of Figure 6 shows the free energy profile of phosphate associations along the committor coordinate. For most of the investigated mutants the free energy decreases with increasing committor value, indicating that binding of phosphate is energetically favorable. Inspecting the free energy profiles of different mutants shows the existence of several minima along the committor coordinate. Such minima indicate that the phosphate ion is more likely to be found at certain positions in space with corresponding committor values and these configurations may be metastable. Interestingly, the two committor iso-surfaces shown in Figure 5 are especially relevant for the phosphate binding process: for each mutant at least one of these two iso-surfaces describes configurations associated with a minimum in its free energy profile. Whenever a minimum could be assigned to one of the iso-surfaces it is marked with a red or blue dot in the free energy profile. Phosphate configurations represented by the outer iso-surface (red) are subsequently termed intermediate 1 and configurations described by the inner iso-surface (blue) are termed intermediate 2.

In the wild-type protein, both intermediate 1 and intermediate 2 free energy minima indicate two metastable configurations of the phosphate before it reaches the binding site. The A197W (see SI) mutant exhibits a very similar profile and an almost equal association rate of 26.4 $M^{-1}s^{-1}$ compared to 27.9 $M^{-1}s^{-1}$ for the wild-type, indicating that this mutation has little effect on the phosphate ion binding capability. For mutants R134Q/K167Q/K175Q (3 mut.) and R134Q/K167Q/K175Q/D21N/D51N/D61N (6 mut.) the intermediate 1 configuration is destabilized and thus only the configurations corresponding to intermediate 2 are found to be metastable. Both mutants have the same three positively charged amino acids replaced by neutral substitutes, but in the 6 mut. mutant the associated loss of charge is compensated by additionally replacing 3 negatively charged amino acids by neutral substitutes. The destabilization of intermediate 1 indicates that residues Arg134, Lys167 and Lys175 are necessary for holding the phosphate ion at the protein surface. Interestingly, losing this kinetic trap along the binding coordinate does not increase the association rate of phosphate: in contrast it is decreased by a factor of 3 for the 6mut. mutant (9.3 $M^{-1}s^{-1}$) and by a factor of ~ 10 for the negatively charged 3mut. mutant

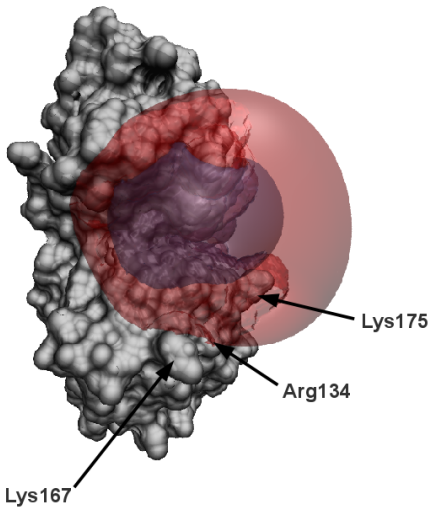

Figure 5: Connolly surface representation of PBP with transparent intermediate 1 (red) and intermediate 2 (blue) committor iso-surfaces.

(2.5 $M^{-1}s^{-1}$). Due to its relevance for attracting phosphates and thereby enhancing the binding efficiency we henceforth abbreviate the positive charge patch around residues Arg134, Lys167 and Lys175 *"anion attractor"*.

To further assess the relevance of positive surface charges, single-point mutations R134Q and R135Q were considered. R134Q neutralizes one residue of the *anion attractor*, whereas R135Q neutralizes a residue which is found between the *anion attractor* and the phosphate binding site, therefore interfering with the phosphate transport route. While both mutants show a reduced association rate, this reduction is 5-fold in R135Q while it is only 2-fold for R134Q. The corresponding free energy profile of R135Q also reveals this effect by showing smaller binding (committor) probabilities for intermediate 1 and 2 configurations than for the R134Q mutant.

The mutants discussed so far have mainly affected residues in the vicinity of the *anion attractor*. For a more comprehensive assessment of phosphate association also mutations D56N, D137T, K43Q and K43M were considered. Mutations D56N and D137T both neutralize a negative charge and increase the association rate by a factor of about 3 compared to the wild-type. Due to thus stronger attraction of the negatively charged phosphate ion, the minimum associated to intermediate 2 configurations vanishes, while intermediate 1 trapping is still present although with an increased probability to reach the binding site from these configurations. The intermediate 2 minimum also disappears for the negatively-charged K43M/K43Q mutants. However, in contrast to the positively charged D56N/D137T mutants, the association rate is reduced by a factor of almost 3 and the binding probability associated with intermediate 1 configurations is strongly reduced, which can be seen from the left shifted minimum in the free energy profiles.

Finally, the T141D mutant is discussed. The free energy profile of this mutant is remarkably different from other mutants that also introduce a negative net charge of -1e. In fact, also a free energy minimum can be assigned to intermediate 1 configurations, but the associated binding probability is very small. Furthermore, the free energy difference between unbound and bound phosphate is positive, rendering phosphate binding unfavorable. This can also be observed, at the corresponding association rate which is also drastically reduced, and a factor of 5 smaller than the smallest association rate found for almost all other mutants with a negative net charge of -1e. An explanation of this result might be the location of the mutation, which introduces a

negative charge very close to the phosphate binding site, repelling the phosphate here. Unlike other mutations that introduce negative charges, in this case the phosphate ion cannot avoid the repulsive region via alternative pathways in order to reach the binding site. Consequently, this mutation has the largest effect on the association efficiency of the phosphate ion.

Table 1: Net charge and computed bimolecular association rates of considered mutant structures at ionic strength of $0mM$.

Mutant	Net Charge $[e]$	k_{on} $[10^8 M^{-1}s^{-1}]$
wt (modeled)	0	27.9
A197W	0	26.4
D56N	+1	73.9
D137T	+1	77.8
T141D	-1	1.6
R135Q	-1	5.9
K43M	-1	11.3
K43Q	-1	11.4
R134Q	-1	12.4
3mut	-3	2.5
6mut	0	9.3
PBP:P_{i}	-2	3.0

4.2 Stream Lines and First Hitting Density

The free energy profiles and rates described above provide information about macroscopic or effective properties of the phosphate association process, but they do provide little information about the fine details of phosphate association. More do the specific dynamical properties that can be accessed with the Transition Path Theory approach: the shape of the binding pathways and the distribution of where they hit the protein surface. Figure 6 and the following show representative pathways of the association pathway ensemble. These plotted pathways are streamlines that follow the reactive flux field of binding. The number of reactive trajectories that pass a volume element per unit of time is expressed by streamline coloring. The brighter the coloring, the more reactive trajectories pass through the surrounding volume elements. This manifests as almost white coloring in the vicinity of the binding site, where the increasing bundling of reactive trajectories leads to an increased flux density. In order to obtain additional information where the phosphate association pathways "attack" the protein, we measured how many reactive trajectories per unit of time hit surface elements in a distance of 10 Å around the protein. This hitting density is visualized by a planar projection in the second column of Figure 6 along with the positions of the mutations.

The neutrally charged structures wt and A197W share a similar pattern in the first hitting density and distribution of pathways. The phosphate trajectories attack the protein on both sides of the phosphate binding side, with a preference for the side at which the *anion attractor* is located. The corresponding stream line illustrations show that some phosphates form first contact with the *anion attractor* and then crawl over the surface to the binding site. This picture is not qualitatively different for the positive D56N and D137T mutants. Here, also both sides of the protein are approached by the phosphate and the surface crawling still occurs. However, due to the increased net charge of the protein the number of reactive trajectories is strongly increased. A change in both hitting density and approach pathway distribution can be observed for in K43M/K43Q. In this case the number of pathways that attack the protein at the side of the mutation is reduced and the stream lines show that the phosphate is no longer attracted to the surface at the respective position. An even stronger distortion is observed when the positive patch is neutralized as in the *6mut.* and *3mut.* mutants. The number of pathways that hit the extended

protein surface above the positive patch is significantly reduced in both cases. Furthermore, the flux lines show that the pathways are not attracted to the positive patch but rather straightly approach the phosphate binding site from the bulk. Due to the negative net charge of the *3mut.* mutant the number of phosphates that reach the binding site is per unit of time is reduced as visible from darker flux lines. While the T141D mutation was found to strongly reduce the association rate it neither exhibits a change in first hitting density nor is the topology of association pathways affected. The surface attraction of the phosphate ion is still present, however, the number of phosphates reaching the binding site is strongly reduced, i.e. this mutation only affects the last step of association. Although mutations R134Q and R135Q do not show a pronounced effect on the first hitting density, they do show a difference in the flux line picture. In comparison to R135Q, the surface attraction at the positive patch is less pronounced in the R134Q mutant.

In the results shown so far, we have investigated the binding dynamics of a single phosphate ion in the dilute limit, i.e. in absence of other solutes. In a biological scenario, the situation is much more complex as the cytosol is densely filled with various species of different sizes, shapes and charges. Although such a heterogeneous complexity is of limited interest to the biophysicist, it is very interesting to work out some of the principles that contribute to the phosphate binding dynamics, and more generally to potentially all ion-binding dynamics, in the cell. For example how does phosphate binding occur in a phosphate rich environment, i.e. where phosphates compete for binding? To model this, we investigate inorganic phosphate association in a model where a phosphate ion is already trapped at the positively charged surface patch. Therefore, a HPO_4^{-2} ion was placed in the vicinity of Arg134, Lys167 and Lys175 and the association dynamics were computed based on the resulting electrostatic potential. The computed free energy profile, the first hitting density and the binding pathways are depicted in Figure 7. The free energy profile shows that the trapping property of the positively charged patch is lost when it is already loaded with a negatively charged ion, the minima corresponding to the intermediate 1 iso-committor surface was not present anymore. Moreover, the overall binding free energy is nearly zero. The hitting density plot shows that the pathways avoid to attack the protein at the bound phosphate location and are redirected further down. The streamlines additionally reveal that the second phosphate does not "crawl" over the *anion attractor*, it rather reaches the binding site from space.

5 Conclusion

In this study we have presented a computational approach to systematically investigate protein-ligand association kinetics. While existing computational approaches have permitted the calculation of binding energies and rates using a variety of molecular and dynamical models, the presented method goes beyond these by providing an extensive analysis of the entire ensemble of association pathways by which a ligand approaches its target protein, and their relative probabilities. While in the present study a simple electrostatic interaction model was used in combination with rigid body BD, our analysis approach can be readily applied to any molecular dynamical model that allows to calculate or estimate transition probabilities or rates between the substates of the protein-ligand configuration space.

The usefulness of the approach was demonstrated by studying the binding of inorganic phosphate to the Phosphate Binding Protein from *Escherichia coli* and several in silico mutants of it. The analysis reveals that protein mutations that affect surface charges may have subtle to drastic effects on the association kinetics and association pathways. Some mutations affect only association rates without significantly altering the associating pathways, i.e., they scale the fluxes. Other mutations change the association pathways of P_i, and the associated change of the rate, may be of very different magnitude depending on the exact location of the mutation.

Overall, all systems studied here exhibit binding via a broad ensemble of parallel pathways, indicating a funnel-like energy landscape that narrows down towards the bound state, very similar to the situation in protein folding (50).

Consequently, only very few single-point mutations are able to effectively disable P_i binding. The only single point mutation observed to do this was next to the binding site and thus affected

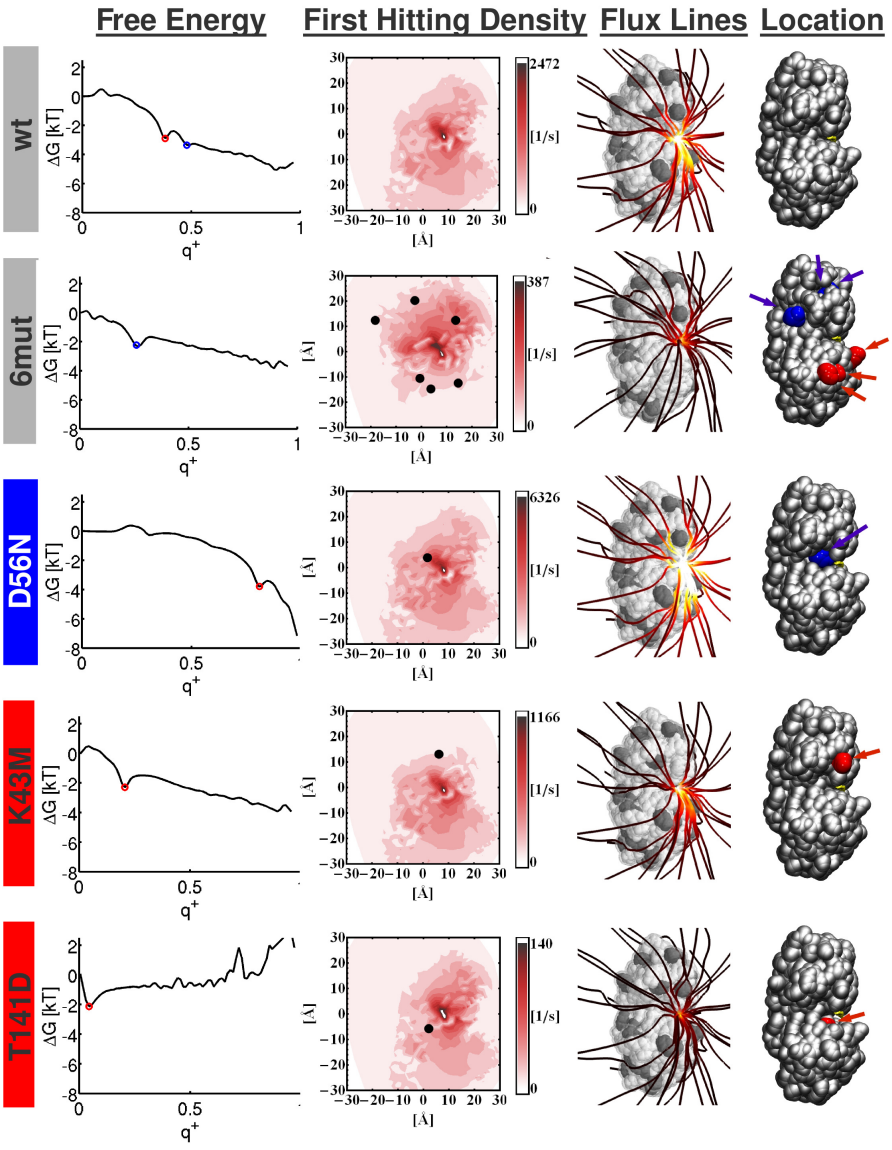

Figure 6: Free energy profiles, first hitting densities, and pathways for selected mutants of P_i associating to the Phosphate Binding Protein. Text background coloring: Gray - Protein has neutral net charge, Blue - positive net charge, Red - negative net charge. *First column* - Free energy profile of the ligand when it travels from the dissociated ($q^+ = 0$) to the associated state ($q^+ = 1$). A red or blue dot is shown whenever a minimum could be assigned to one of the two iso-surfaces shown in Figure 5. *Second column* - Surface density of reactive trajectories that hit the extended protein surface per unit of time. In each of the plots the black points represent projected C_α positions of mutated amino acids. Note that the color axis is scaled separately for each mutant for clarity. *Third column* - Streamlines of the reactive flux of ligand association (see Sec. 3.6) for the different mutants, which represent the ensemble of association pathways. A lighter streamline coloring corresponds to a higher local reactive flux. Here, the same color scheme is used for all pictures. *Fourth column* - Here the positions of mutated residues are shown. Red/blue corresponds to negative/positive charged mutations relative to the wild-type structure.

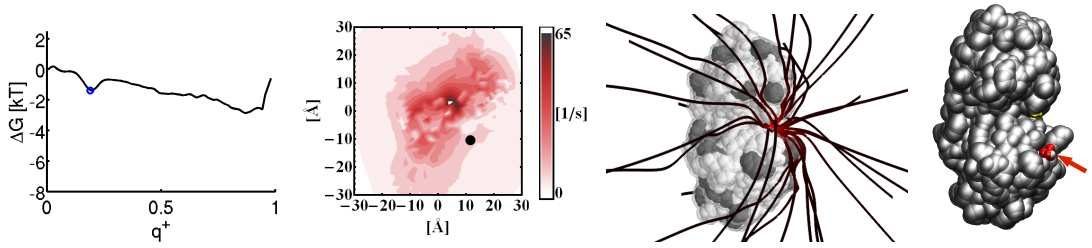

Figure 7: Phosphate Binding to PBP:P_i - free energy profile, first hitting density, association pathways, marked phosphate location

nearly all binding pathways at the bottleneck where they converge. Most other constructed single-point mutations only disabled a subset of pathways, allowing other parts of the pathway ensemble to take over, resulting in only a mild reduction of the association rate. Multiple mutations at critical positions, however, were much more effective and could efficiently disable binding.

The analysis of the mutagenetic behavior revealed the importance of two anion attractors on the surface of PBP which unspecifically attract all negatively charged molecules. This unspecific attraction brings anions closer to the phosphate binding site thus trapping them in a region of limited size. As a result, the PBP wild-type exhibits "superdiffusive association", i.e., associations with a rate that is about three fold compared to the free-diffusion association rate to the binding site that is estimated to be 9.2 $M^{-1}s^{-1}$. With favorable mutations, the association rate may be sped up to about ten times the free diffusion rate.

After an anion reaches the attractor the phosphate specificity is introduced in the subsequent step, i.e., the actual complex formation, where binding of phosphate is energetically favorable over other anions by detailed interaction (25, 27). The resulting catch and select mechanism might be a general strategy which allows ions to be efficiently screened before being specifically selected. To experimentally verify our findings, the relevance of different pathways on the protein surface might be assessed by labeling of specific surface residues and P_i and investigation of their contact dynamics using, e.g., NMR.

The grid based discretization used here to define configurational substates is limited to few dimensions and is thus limited to study systems of a size like ligand approaching a rigid protein. However, in future work the approach will be extended to gridless data-based discretization of configuration spaces as they are frequently used in Markov model analysis of protein internal dynamics (15). With this extension, a flux analysis of association pathways will be possible for complex protein-ligand and protein-protein binding with full dynamical treatment such as all-atom MD in explicit solvent.

6 Acknowledgments

MH and FN acknowledge financial support by the DFG SFB 449 and the IMPRS-CBSC. PM, JHP and FN acknowledge financial support by the DFG Research Center MATHEON.

References

1. Ostermann, A., R. Waschipky, F. G. Parak, and U. G. Nienhaus, 2000. Ligand binding and conformational motions in myoglobin. *Nature* 404:205–208.

2. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes, 1991. The energy landscapes and motions of proteins. *Science* 254:1598–1603.

3. Tummino, P. J., and R. A. Copeland, 2008. Residence time of receptor-ligand complexes and its effect on biological function. *Biochemistry* 47:5481–5492.

4. Csermely, P., R. Palotai, and R. Nussinov, 2010. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochem. Sci.* .

5. Lange, O. F., N.-A. Lakomek, C. Fares, G. F. Schroder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger, and B. L. de Groot, 2008. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320:1471–1475.

6. O'Toole, N., and I. A. Vakser, 2008. Large-scale characteristics of the energy landscape in protein-protein interactions. *Proteins* 71:144–152.

7. F. Noé and S. Doose and I. Daidone and M. Löllmann and J.D. Chodera and M. Sauer and J.C. Smith, 2010. Dynamical Fingerprints: Understanding biomolecular processes in microscopic detail by combination of spectroscopy, simulation and theory. *submitted to Proc. Natl. Acad. Sci. USA on 20 Apr 2010. In revision* .

8. Schluttig, J., D. Alamanova, V. Helms, and U. S. Schwarz, 2008. Dynamics of protein-protein encounter: A Langevin equation approach with reaction patches. *J. Chem. Phys.* 129:155106.

9. Northrup, S. H., and H. P. Erickson, 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA* 89:3338–3342.

10. Northrup, S. H., S. A. Allison, and J. A. McCammon, 1984. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80:1517–1524.

11. Gabdoulline, R. R., and R. C. Wade, 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.

12. Gabdoulline, R. R., and R. C. Wade, 2001. Protein-Protein Asociation: Investigation of factors influencing association rates by Brownian Dynamics Simulation. *J. Mol. Biol.* 306:1139–1155.

13. Spaar, A., C. Dammer, R. R. Gabdoulline, R. C. Wade, and V. Helms, 2006. Diffusional Encounter of Barnase and Barstar. *Biophys. J.* 90:1913–1924.

14. McGuffee, S. R., and A. H. Elcock, 2010. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput. Biol.* 6:e1000694.

15. Noé, F., C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, 2009. Constructing the Full Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. USA* 106:19011–19016.

16. Noé, F., 2008. Probability Distributions of Molecular Observables computed from Markov Models. *J. Chem. Phys.* 128:244103.

17. Chodera, J. D., K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, 2007. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101.

18. Bowman, G. R., K. A. Beauchamp, G. Boxer, and V. S. Pande, 2009. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* 131:124101.

19. Buchete, N. V., and G. Hummer, 2008. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* 112:6057–6069.

20. Song, Y., Y. Zhang, T. Shen, C. L. Bajaj, A. A. McCammon, and N. A. Baker, 2004. Finite Element Solution of the Steady-State Smoluchowski Equation for Rate Constant Calculations. *Biophys. J.* 86:2017–2029.

21. Kube, S., and M. Weber, 2007. A coarse graining method for the identification of transition rates between molecular conformations. *J. Chem. Phys.* 126:024103.

22. Prinz, J.-H., H. Wu, M. Sarich, B. Keller, M. Fischbach, M. Held, J. D. Chodera, C. Schütte, and F. Noé, 2010. Markov models of molecular kinetics: Generation and Validation. *submitted to J. Comput. Phys.* .

23. Metzner, P., C. Schütte, and E. Vanden-Eijnden, 2006. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* 125.

24. Schütte, C., F. Noé, E. Meerbach, P. Metzner, and C. Hartmann, 2009. Conformation Dynamics. *In* R. Jeltsch, and G. W. (Eds), editors, Proceedings of the International Congress on Industrial and Applied Mathematics (ICIAM). EMS publishing house, 297–336.

25. Yao, N., P. S. Ledvina, A. Choudhary, and F. A. Quiocho, 1996. Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. *Biochemistry* 35:2079–2085.

26. Wang, Z., A. Choudhary, P. S. Ledvina, and F. A. Quiocho, 1994. Fine tuning the specificity of the periplasmic phosphate transport receptor. Site-directed mutagenesis, ligand binding, and crystallographic studies. *J. Biol. Chem.* 269:25091–25094.

27. Luecke, H., and F. A. Quiocho, 1990. High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature* 347:402–406.

28. Brune, M., J. L. Hunter, S. A. Howell, S. R. Martin, T. L. Hazlett, J. E. T. Corrie, and M. R. Webb, 1998. Mechanism of Inorganic Phosphate Interaction with Phosphate Binding Protein from Escherichia coli. *Biochemistry* 37:10370–10380.

29. Ledvina, P. S., A. L. Tsai, Z. Wang, E. Koehl, and F. A. Quiocho, 1998. Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Sci.* 7:2550–2559.

30. Huang, H.-C., and J. M. Briggs, 2002. The association between a negatively charged ligand and the electronegative binding pocket of its receptor. *Biopolymers* 63:247–260.

31. Gabdoulline, R. R., and R. C. Wade, 1999. On the protein-protein diffusional encounter complex. *J. Mol. Recognit.* 12:226–234.

32. Geyer, T., and U. Winter, 2009. An O(N^2) approximation for hydrodynamic interactions in Brownian dynamics simulations. *J. Chem. Phys.* 130:114905.

33. Fogolari, F., A. Brigo, and H. Molinari, 2002. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* 15:377–392.

34. Weinan, E., and E. vanden Eijnden, 2006. Towards a Theory of Transition Paths. *J. Stat. Phys.* 123:503–523.

35. Metzner, P., C. Schütte, and E. vanden Eijnden, 2009. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* 7:1192–1219.

36. Metzner, P., C. Schütte, and E. V. Eijnden, 2009. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* 7:1192–1219.

37. and, R. E., 2009. Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.* 6:046001+.

38. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

39. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, 2005. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

40. MacKerell, A. D., D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* 102:3586–3616.

41. Li, H., A. D. Robertson, and J. H. Jensen, 2005. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61:704–721.

42. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker, 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665–667.

43. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98:10037–10041.

44. Kielman, H. S., and J. C. Leyte, 1975. Selfdiffusion of phosphate and polyphosphate anions in aqueous solution. *In* in: Magnetic Resonance, and R. Phenomena, editors, Proceedings of Congress AMPERE. volume 2, 515–516.

45. Latorre, J., P. Metzner, C. Hartmann, and C. Schütte, 2010. A Structure-preserving numerical discretization of reversible diffusions. *submitted to Comm. Math. Sci.* .

46. Matrix Toolkits Java - http://code.google.com/p/matrix-toolkits-java/.

47. J.-H. Prinz and M. Held and J. C. Smith and F. Noé, 2009. Efficient Computation of Committor Probabilities and Transition State Ensembles. *submitted to SIAM Multiscale Model. Simul.* .

48. Du, R., V. S. Pande, A. Yu, T. Tanaka, and E. S. Shakhnovich, 1998. On the transition coordinate for protein folding. *J. Chem. Phys.* 108:334–350.

49. Ramachandran, P., and G. Varoquaux, 2008. Mayavi: Making 3D Data Visualization Reusable. *In* G. Varoquaux, T. Vaught, and J. Millman, editors, Proceedings of the 7th Python in Science Conference. Pasadena, CA USA, 51–56.

50. Dill, K. A., and H. S. Chan, 1997. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* 4:10–19.

7 Supplement

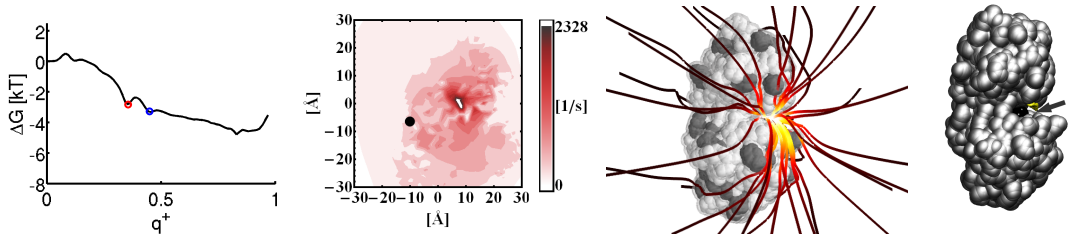

Figure 8: Mutant A197W - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site

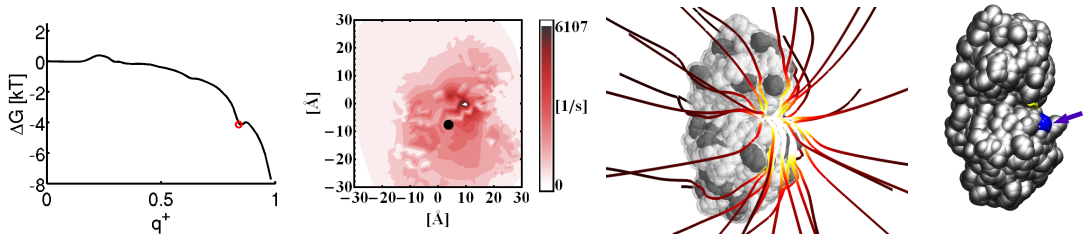

Figure 9: Mutant D137T - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site

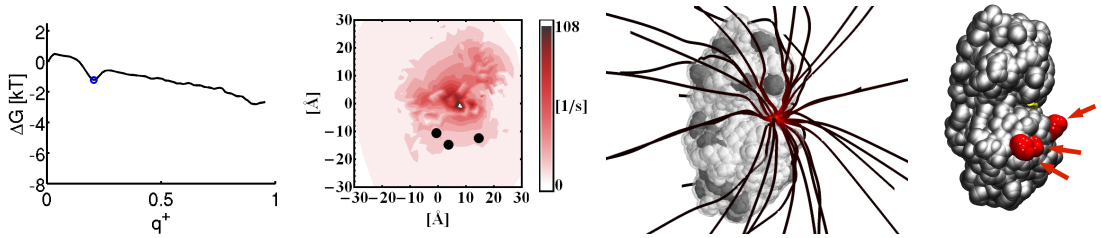

Figure 10: Mutant 3 mut. - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site

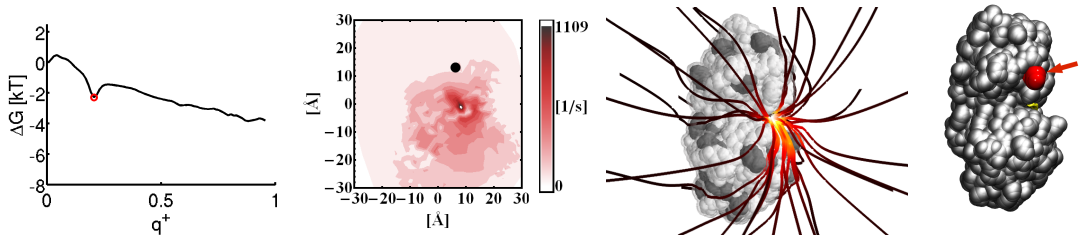

Figure 11: Mutant K43Q. - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site

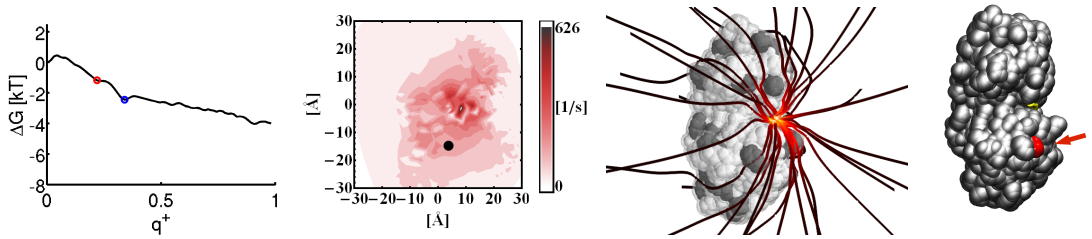

Figure 12: Mutant R134Q. - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site

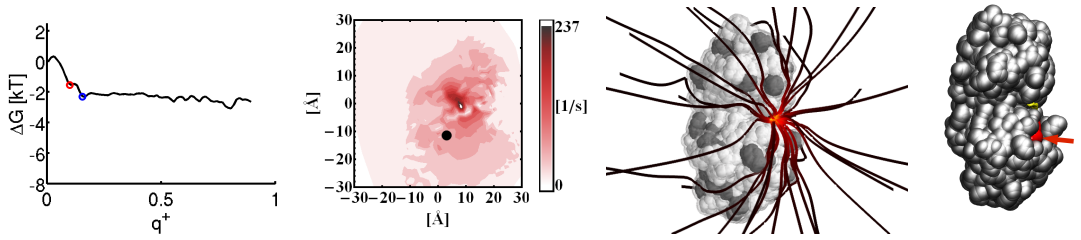

Figure 13: Mutant R135Q. - Committor Free Energy Profile, First Hitting Density, Association Pathways, Mutation Site