Calculating Kinetics and Pathways of Protein-Ligand Association

published in EJCB, 91 (2012), pp. 357-364

Martin Held[1] and Frank Noé[1]

[1]Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

February 23, 2012

Abstract

While studies of protein-ligand association have mostly focused on the native complex and its stability (binding affinity), relatively little attention has been paid on the association process that precedes the formation of the complex. Here we review approaches to study the kinetics of association and association mechanisms, i.e. the probability distribution of association pathways. Selected methods are described that allow these properties to be calculated quantitatively from simulation models. We summarize some applications of these methods and finally propose a model mechanism by which proteins may efficiently screen potential ligands for those that can be natively bound.

Protein-protein and protein-ligand interactions are essential for almost any biological process. Protein-ligand and protein-substrate binding is a prerequesite for signal transduction and modulation processes. Protein-protein interactions are for example required to form filaments that are stabilized by strong interactions amongst the constituing monomers. Due to their importance protein interactions have been subject to intensive investigation since the early days of molecular biology [Pauling et al., 1943; Kauzmann, 1959]. Classically, protein interactions have largely been characterized in terms of the affinity of the interaction partners involved. For example, a highly affine interaction is characterized by the fact that small concentrations of the respective interaction partners lead to a high yield of interaction complex. While affinity has been proven to be a good measure to indicate for example the potency of an enzyme inhibitor it "hides" the kinetics of the underlying structural and biochemical processes. This kinetic aspect can be crucial to assess the effectivity of a certain reaction [Tummino and Copeland, 2008] or to foster a biophysical understanding of the reaction process itself. For example, certain interactions show a high affinity due to a slow k_{off} rate, which means the interaction partners form a complex that is stable for a rather long time, while other interactions might exhibit the same affinity value while with a higher turnover, i.e., a faster association rate (k_{on}) and also a faster dissociation rate (k_{off}). A prominent example for the biological relevance of kinetically fast interactions is given by the degradation of the neurotransmitter Achethycholine (ACh) from the synaptic cleft, which is carried out by the enzyme Achetylcholinesterase (AChE). For this process very fast kinetics are indispensable, as fast high frequency transmission of nerval signals relies on a quick restoration of the transmitting modules. As a result of this requirement, evolution made AChE to one of the fastest known enzymes, with a protein-ligand association that is super-diffusive, i.e. faster than would be realized by free diffusion [Ripoll et al., 1993; Radic, 1997; Meltzer et al., 2006].

In the following we give a short overview of theoretical studies that have contributed to investigate mechanisms and kinetics of protein interactions and present a new methodology that in principle allows to compute the ensemble of all reactive protein-protein/protein-ligand association pathways. This approach is illustrated on an example of binding of a phosphate ion to the Phosphate Binding Protein, previously published in [Held et al., 2011]. For a more thorough review of existing theoretical work on protein-protein interactions the reader is referred to previous studies, e.g., [Vijayakumar et al., 1998; Gabdoulline and Wade, 2002; Papoian and Wolynes, 2003; Dell'Orco, 2009; Schreiber et al., 2009].

In many studies, protein association dynamics are modelled as a random diffusion process in an energy landscape that arises from the electrostatic interaction energy of the proteins involved [Northrup and Erickson, 1992; Zhou, 1993; Gabdoulline and Wade, 1997a; Ehrlich et al., 1997; Gabdoulline and Wade, 1999; Elcock et al., 1999; Altobelli and Subramaniam, 2000; Gabdoulline and Wade, 2001a; Elcock, 2003; Haddadian and Gross, 2006; McGuffee and Elcock, 2010; Pachov et al., 2011]. In these approaches the influence of individual water molecules in the system is approximated via a continuum heat bath that exerts random kicks on the solute molecules. Furthermore, it is often assumed that the internal degrees of freedom, e.g., flexibility of side-chains or internal conformational changes can be neglected and the proteins are thus modeled as rigid bodies. These simplifications allow to reach simulation timescales that are relevant for protein association events at a comparatively low computational cost compared to all-atom molecular dynamics. It should be noted that with advances in computational power, such fully detailed molecular dynamics simulations to study the association process are becoming accessible [Ahmad et al., 2008; Buch et al., 2011]. Protein flexibility is likely to be essential for the actual binding process that occurs after association

towards a pre-complex, either because the binding induces conformational changes, or because it proceeds via conformational selection [Lange et al., 2008].

In 1978 Ermak and McCammon [Ermak and McCammon, 1978] have presented an algorithm that allows mixtures of proteins to be simulated with Brownian dynamics. Based on this algorithm, a strategy has been developed to compute bi-molecular kinetic association constants [Northrup et al., 1984]. To obtain the association rate of two proteins the procedure is the following: One protein is translationally constrained in the center of a sphere and the other is placed in random orientations at random position at the surface of the sphere. Starting from these configurations the Ermak-McCammon algorithm is used to simulate the movement of the diffusing protein for each of the generated setups. Based on the obtained simulation results the fraction of proteins that successfully associated with the center protein is estimated. This fraction together with the expected Smoluchowski rate to find both proteins at the radius of the sphere from where the trajectories started is then used to compute the diffusional association rate. This approach to calculate a bi-molecular association rate constant is often abbreviated as NAM (Northrup, Allision, McCammon) approach. In the initial [Northrup et al., 1984] publication this rate was based on branching diagrams, but later an analytical expression has been derived by Zhou [Zhou, 1990].

In the past 25 years the described methodology was very successfully applied to study the mechanisms and kinetics of a number of biological interactions processes. Ranging from studies of bi-molecular protein-protein interactions [Northrup and Erickson, 1992; Zhou, 1993; Gabdoulline and Wade, 1997a; Elcock et al., 1999; Gabdoulline and Wade, 1999; Altobelli and Subramaniam, 2000; Gabdoulline and Wade, 2001a; Haddadian and Gross, 2006], over DNA-chromatin [Ehrlich et al., 1997; Pachov et al., 2011] interactions to studies of heterogeneous protein mixtures Elcock [2003]; McGuffee and Elcock [2010]. In parallel to the complexity increase of the systems studied with Brownian dynamics algorithms, also the underlying models improved. For example, the different dielectric nature of protein and solvent had been neglected in early works. Gabdoulline and Wade [Gabdoulline and Wade, 1996] have developed a the effective charge method that elegantly accounts for this difference in an approximate but computationally efficient way. Furthermore, hydrodynamic interactions had been neglected in Brownian dynamics studies due to their small *accuracy benefit* vs. *computational cost* ratio. However, recent algorithmic advances by Geyer et al. [Geyer and Winter, 2009] substantially reduce the associated computational cost and render their future inclusion in BD simulations economic.

Despite the success of BD to correctly model association kinetics, little work was devoted to systematically characterize the association and dissociation pathways that lead to formation of protein complexes, i.e. the microscopic mechanism. An exception in that regard is the work by Spaar et al. [Spaar et al., 2006b], who analysed the fraction of trajectories that successfully formed an Barnase-Barstar encounter complex. While their analysis provides valuable insights about possible association and dissociation pathways, it remains largely qualitative. It is for example not apparent from their analysis which pathways connect the metastable regions present in the vicinity of the Barnase protein.

To overcome this lack of quantitative analysis, we have recently developed an approach that allows to compute both the association kinetics and the ensemble of association pathways of protein-ligand association processes in a rigorous way. This approach is based on Transition Path Theory, which is a statistical physics approach to characterize transition pathways between selected subsets of a dynamical system. In the remainder of this paper this approach will be recapitulated and illustrated at the example of inorganic phosphate binding to the

Phosphate Binding Protein. Further we propose a kinetic model that models the binding of anions to the attractive patch of the Phosphate Binding Protein, revealing a potential kinetic benefit of low-affinity binders over high-affinity binders.

Calculation of transition pathways and rates

How can the full ensemble of pathways and the overall rate of a protein-ligand association process be calculated? First, a molecular model needs to be defined, i.e. the computational representation of the protein, its ligand, and their interactions with each other and with the solvent. At very small interaction distances structural dynamics and explicit water molecules are important. Therefore, explicit water all-atom molecular dynamics simulations have been used to investigate questions such as whether ligand binding occurs via induced fit or conformational selection [Lange et al., 2008]. For binding partners that have relatively stable structures and at distances where they are still separated by one or two solvation layers, the internal dynamics of protein and ligand and the explicit structure of solvent molecules are presumably less important. Consequently, it is common at this stage to use models that keep the internal structures of protein and ligand rigid and treat the solvating environment (consisting of water, ions, and other non-resolved molecules) implicitly by subsuming it into a (1) an effective diffusion constant, governing how fast protein and ligand can move through this media, and (2) an effective dielectric permittivity, governing the strength of electrostatic interactions between protein and ligand, potentially also accounting for the effect of the ionic strength of the solution. Based on these two properties and the specific location of charges on protein and ligand, their motion can be described by Brownian dynamics (BD) [Northrup et al., 1984; Gabdoulline and Wade, 1997b, 2001b; Spaar et al., 2006b; Schluttig et al., 2008]. Having defined the molecular model and dynamics, all properties of interest are in principle defined, including the binding affinity, the binding rate, and the ensemble of binding pathways. However, it is often impractical to calculate these properties by calculating time averages from direct simulation, i.e. by running a single or a number of long trajectory/-ies that bind/s and unbind/s sufficiently often.

Recently, mathematical and computational approaches have been developed that break down this problem into sub-problems each of which are practically manageable. We sketch the approach here and give the most important expressions for rates and pathway probabilities. First, the conformation space of protein and ligand are conceptually divided into the dissociated regime, where the protein-ligand pair under investigation do not interact and diffuse freely through the media, and the near regime where one protein copy and one ligand copy are considered to be close enough such that they interact, e.g. *via* electrostatic and hydrodynamic forces (Fig. 1). The near regime is entered with a rate k_{Diff} that depends on the concentration of protein and ligand, and the ligand can then either diffuse out again or associate with the protein at a rate k_{AB}. k_{AB} is the rate of a complex formation event of a single protein-ligand pair from the boundary of the near regime, i.e. it is measured in s^{-1} and is not concentration dependent, and it depends on all details of the interaction between protein and ligand.

While k_{Diff} can be easily calculated from the Smoluchowski rate equation (see Fig. 1) the calculation of the ensemble of pathways in the near regime and its associated rate k_{AB} is more difficult. The key here is to subdivide this near regime into small conformational substates, calculate the transition rates or transition probabilities between pairs of such substates i and

j separately, and then reassemble the entire dynamics within the near regime from it. We and others have developed and established such *ensemble dynamics* or *Markov models* in the past few years [Schütte et al., 1999; Swope et al., 2004; Noé et al., 2007; Chodera et al., 2007; Noé, 2008; Noé and Fischer, 2008; Buchete and Hummer, 2008; Noé et al., 2009; Bowman et al., 2009; Voelz et al., 2010; Prinz et al., 2011,(; Keller et al., 2011]. The typical application of Markov models is to analyze complex processes such as protein folding where large amounts of molecular simulation is clustered into small conformational substates and the overall process can be understood by analyzing the resulting transition probabilities between these substates [Karpen et al., 1993; de Groot et al., 2001; Weber, 2003; Wales, 2003; Rao and Caflisch, 2004; Schultheis et al., 2005; Hubner et al., 2006; Muff and Caflisch, 2007; Noé and Fischer, 2008; Buchete and Hummer, 2008; Pan and Roux, 2008]. The dynamical data may be generated with any dynamical model, including all-atom molecular dynamics, coarse-grained dynamics or Brownian dynamics with rigid solutes. Such analyses are especially useful in scenarios where large simulation datasets arise, such as in the folding@home framework [Singhal et al., 2004; Bowman et al., 2009]. On the other hand, in situtations such as protein-protein or protein-ligand binding, essential degrees or freedom are often known *a priori*. For example, obviously relevant coordinates are the 6 rototranslational degrees of freedom that describe the relative position and orientation of the second protein or ligand with respect to the first protein. In such cases, the conformation space can be discretized a priori and many short simulations can be started from many different substates, without having to wait for the dynamics to sample these substates. One advantage of Markov models over standard analyses is then that their substate-to-substate transition probabilities $T_{ij}(\tau)$ can be estimated from many short trajectories of length τ that may be much shorter than the slowest relaxation times of the system, thus getting somewhat independent of the rare event character of the molecular system . [Noé et al., 2009; Prinz et al., 2011]. This approach introduces statistical uncertainty in the results whose magnitude can be estimated with novel statistical methods [Singhal and Pande, 2005; Hinrichs and Pande, 2007; Noé, 2008; Chodera and Noé, 2010].

In simple cases, such as the diffusive binding of an ion to a rigid protein, the conformational substates can be defined by a three-dimensional grid in the space of ion location around the protein. In this case, there are even no simulation trajectories needed because it is possible to calculate transition rates k_{ij} between substates from a direct spatial discretization of the dynamical equations and without statistical error [Song et al., 2004; Held et al., 2011]. Markov models can be expressed in terms of either transition probabilities $T_{ij}(\tau)$ or transition rates k_{ij}, and the subsequent analysis also works with both approaches. For simplicity, we continue with the transition rates k_{ij}.

A number of quantities can be calculated from the transition rates k_{ij}. Each conformational sub-state has a stationary distribution π_i (the Boltzmann distribution, see Fig. 2b) for which the detailed balance equation $\pi_i k_{ij} = \pi_j k_{ji}$ holds, i.e. in equilibrium, the flux of trajectories from state i to j, $\pi_i k_{ij}$, equals the flux of trajectories backward, $\pi_j k_{ji}$.

We now want to concentrate on those trajectories only that come from the dissociated state A (i.e. the boundary between far and near regime) and progress to the associated state B without returning to A in between. We are interested in characterizing the probability distribution of these *reactive trajectories*, and in the flux these trajectories generate between the state. For this, we require the concept of the committor q_i, which is, for each state i, the probability that a trajectory being in this state will next move on to associate (to B), rather than dissociate (to A) [Du et al., 1998; Bolhuis et al., 2002; Vanden-Eijnden, 2006], see Fig. 2 for illustration. The committor can be easily calculated from k_{ij} [Noé et al., 2009].

As q_i quantifies the kinetic progress of the ligand association process, it can be used as a reaction coordinate or order parameter for a free energy profile. The free energy profile of ligand association is then given by:

$$F(q^*) = -k_{\mathrm{B}}T \log \sum_{i|q_i \approx q^*} \pi_i, \tag{1}$$

where k_{B} is the Boltzmann constant, T the absolute temperature and q^* the value of the committor coordinate. More importantly, the committors can be used to calculate the reactive fluxes f_{ij}^+ between substates via transition path theory (TPT) [Vanden-Eijnden, 2006; Metzner et al., 2009]. The reactive flux is the number of trajectories passing from i to j per time unit moving from the dissociated to the associated state, and is given (at the equilibrium) by:

$$f_{ij}^+ \quad = \quad \max\{\pi_i k_{ij}(q_j - q_i), 0\}. \tag{2}$$

which defines a flux field that can be used to show the streamlines out of A and into B along with their probabilities (see Fig. 2d).

Finally the $A \to B$ reaction rate is given by [Noé et al., 2009]:

$$k_{AB} = \frac{\sum_{i \in A} \sum_{j \notin A} \pi_i k_{ij} q_j}{\sum_i \pi_i (1 - q_i)}.$$

k_{AB} is the rate at which a ligand molecule binds starting from set A. In order to compute the bi-molecular association rate of protein and ligand, 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 with diffusion constant D outside the A sphere, the diffusion limited association constant k_{On} can be obtained by [Erban and Chapman, 2009]:

$$k_{\mathrm{On}} = 4\pi D \left(r - \sqrt{\frac{D}{k_{AB}}} \tanh \left(r \sqrt{\frac{k_{AB}}{D}} \right) \right), \tag{3}$$

where r denotes the radius of the A sphere. Note that k_{On} is a concentration dependent rate (e.g. in $M^{-1}s^{-1}$), while k_{AB} is the rate of a single complex binding event (in s^{-1}). In the derivation of this formula Erban and Chapman assume that molecules X diffuse to the surface of a sphere of radius r, which is centered around molecule Y. Once molecules X hit the surface of the sphere they are removed from the surface with rate k_{AB}, i.e. the rate at which ligands bind given they are found on the surface of a sphere with radius r. The approach described here is an alternative to the NAM approach [Northrup et al., 1984] to calculate binding rates k_{On}. Both approaches converge to the exact binding rates when statistical and discretization errors approach zero. Our approach has the advantage that it can deal with substate-based dynamical models such as Markov models that can be generated even for processes that are much slower than affordable simulation lengths.

Phosphate Binding to the Phosphate Binding Protein

A number of studies, especially using Brownian dynamics simulations of rigid proteins and ligands, have been used to study protein-ligand association kinetics [Gabdoulline and Wade,

1997c, 2001b; Spaar et al., 2006a, 2009; Held et al., 2011]. These studies have found relatively broad ensembles of association pathways that narrow down when the ligand approaches the binding site. In some systems, the dependence of these pathways on mutations were investigated [Spaar et al., 2006a, 2009; Held et al., 2011]. Examples have been found that association might occur via *pre-binding sites* outside the native binding site that metastably associate ligands, and from which either binding or dissociation might occur. In [Held et al., 2011], the binding of phosphate (P_i) to the Phosphate Binding Protein was studied in detail using the systematic approach described in the above section. The rate matrices K, which served as input for the TPT analysis, were computed by discretizing the surrounding space of the Phosphate Binding Protein mutants into a number of volume elements. Given this discretization the transition rates k_{ij} between volume elements were calculated based on the joint diffusion constant of protein and ligand and the electrostatic potential associated to each volume element. The unbound set of states A was chosen to comprise all elements with a distance larger then 25nm to the protein center, the set of bound states B was chosen to comprise elements in the vicinity of the bound P_i conformation, please see [Held et al., 2011] and [Latorre et al., 2010] for details. While previous work on this system was mainly concerned with investigating the binding kinetics by experimental means [Brune et al., 1998; Ledvina et al., 1998] or direct simulation [Huang and Briggs, 2002], our study was the first to provide a systematic description of the P_i binding pathway ensemble.

Similar to diffusion-controlled enzymes [Stroppolo et al., 2001] we find that the wild-type shows association rates greater than the diffusion-limited rate which would be expected for finding the binding site by pure undirected diffusion. By investigating several in-silico mutants of the phosphate binding protein this super diffusive binding effect could be attributed to a positively charged patch ("anion attractor") which strongly attracts anions by electrostatic steering. For the purpose of this review, we pick four of the studied proteins that are most instructive to explain the general biophysical mechanisms that alter the binding kinetics and association pathway ensemble.

The results are summarized in Fig. 3. The left column shows the free energy profile of phosphate associations along the committor coordinate and the calculated association rate. For the selected mutants the free energy decreases with increasing committor value, indicating that binding of phosphate is energetically favorable. Inspection of 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.

In particular, the free energy profile of the wild-type protein shows two free energy minima that indicate two metastable configurations of the phosphate before it reaches the binding site. While this indicates a kinetic trapping of phosphate ions before they reach the binding site the calculated association rate of 27.9 $M^{-1}s^{-1}$ is still three times greater than the pure diffusion rate. When the positive patch formed by R134, K167 and K175 is neutralized, as in the R134Q/K167Q/K175Q/D21N/D51N/D61N (6 mut.) mutant, the anion attractor is disruped. Interestingly, loosing 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 (9.3 $M^{-1}s^{-1}$), now being similar to the purely diffusional association rate.

In the mutant D137T a negative charge close to the binding site is neutralized, causing the opposite effect: The association rate is increased by a factor of about 3 compared to the wild-type. Due to this stronger attraction of the negatively charged phosphate ion, the minimum associated to trapped configurations close to the binding site vanishes, while trapping at the

anion attractor is still present, although with an increased probability to reach the binding site from these configurations.

In order to assess the association mechanism, the ensembles of association pathways are shown in the middle column of Fig. 3. The 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, i.e. brighter color means more flux. This manifests as a nearly white coloring in the vicinity of the binding site, where the trajectories converge to form a strong current.

For the wild-type structure, the phosphate trajectories attack the protein on both sides of the phosphate binding side, with a preference for attacking at the *anion attractor* and then crawling over the surface to the binding site. This picture is not qualitatively different for the positively charged 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 strong distortion is observed when the positive patch is neutralized as in the *6mut.* mutant. 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.

In the results shown so far, we have investigated the association dynamics of a single P_i 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. Therefore, 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 when it competes with other anions or other phosphates? 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 and the binding pathways are depicted in the last row of Figure 3. 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 first trapping configurations was not present anymore. Moreover, the overall binding free energy is nearly zero and the binding rate is strongly reduced. The streamlines additionally reveal that the second phosphate does not "crawl" *via* the *anion attractor*, it rather reaches the binding site from bulk solvent.

Rapid scanning of ligands

In vivo, the cell is densely filled with molecules of all sorts, including proteins, ligands, water, ions, RNA, etc. Most complexes, when formed, however, are very specific, e.g. a particular protein will be able to bind one particular ligand or a small class of ligands, but certainly not every ligand that has roughly the right size, shape and overall charge. This is because tight binding is specific in many sites, i.e. formation of hydrogen bonds, electrostatic complementarity or match of hydrophobic patches need to be favorable enough to overcome the associated entropic cost.

Let us remind the reader to the comparison to protein folding, where the entropically favorable unfolded chain is stabilized in a compact state by native interactions. In the early

days of protein folding, Levinthal raised the so-called *Levinthal paradox*, which referred to the kinetics of protein folding: Levinthal assumed that each amino acid could assume at least two conformations (*via* its Backbone rotameric flexibility), all being equally probable except for the stable native state. If it then takes a certain waiting time (e.g. pico- or nanoseconds) to try one such combination, then, given the typical number of amino acids in the protein, so many trials would be necessary that the protein could not fold in the age of the universe. The resolution of this paradox is that different non-native states are *not* equally likely, but tend to become increasingly likely as the native state is approached - which has often been modeled by a protein folding funnel [Thirumalai and Woodson, 1996; Dill, 1999; Onuchic and Wolynes, 2004].

A similar problem seems to appear in protein-ligand binding. When each binding attempt of a wrong ligand to a protein takes a certain time (*e.g.* nano- or microseconds), can then a correct ligand be found within a reasonable timescale at all unless the concentration is very high? It is very likely that cells have developed efficient sorting and searching mechanisms such that this process is not governed by random trial alone. One aspect is certainly that proteins with related binding partners are often located in proximal positions in the cell. However, there might be additional mechanisms that guide binding partners to attract candidates and then rather quickly reject mismatches.

Consider the Phosphate Binding Protein discussed above in a mixture of anions, its ligand P_i at concentration c_L and other non-ligand anions with the same polarity at concentration c_0. The protein will attract all anions to its pre-binding site, the anion attractor, which is a quick way of screening anions, which are at least known to be in the same charge class in which the ligand also is in. Being at the pre-binding site, the non-ligand (0) cannot bind, and must dissociate some time later, while the ligand (L) can bind tightly. The reaction diagram below describes a kinetic model for this scenario:

Based on this reaction diagram, we can calculate the mean time needed to bind a ligand molecule from the bulk, depending on the relative concentration $c_L : c_0$ and the pre-binding affinity $K_a = k_{on}/k_{off}$ with k_{on} being given by Eq. 3. The results shown in Fig. 5 suggest that a high pre-binding affinity is optimal in the case where no non-ligand are around and thus each binding attempt results in success. In this case, increasing the pre-binding affinity reduces the expected time needed for binding as the associated state PL is then more likely to move on to the bound state PL^* rather than to dissociate. However, the picture drastically changes as soon as non-ligands are in bulk. While it is still true that the time needed for a trial with the true ligand is reduced, this is not the case for non-ligands. Whenever a non-ligand is associated, a high pre-binding affinity will force this non-ligand to stay at the pre-binding site for a long time before it can dissociate and free the site for the next trial. The optimal settings in this case, i.e. the situation with minimal time needed to find and bind the ligand, is given for relatively small pre-binding affinities. This illustrates how decreasing the pre-binding affinity can be favourable to speed up binding in the sense that the waiting time to the next successful binding event is kept small since blocking times from „wrong“ ligands are avoided.

Discussion

While protein-ligand binding has in the past mostly focused on measurement or calculation of binding affinities, the field is now clearly moving more towards kinetics and mechanisms. This is in part due to improved experimental techniques, especially single-molecule experiments

such as force probe measurements or single molecule fluorescence, which allow the existence of multiple metastable states and the transition rates between them to be observable. On the theoretical side, methods in the classes of transition networks / Markov models have been developed that allow simulations to be analyzed and interpreted in a way that yields rather unambiguous access to the relevant states, interstate transitions rates and other kinetic quantities. With these methods available, kinetic and mechanistic properties are likely to be in the focus of future research in protein-ligand binding.

One of the issues that have been raised here is the question how a protein can efficiently find its specific binding partner(s) in a vast and dense mixture of different possible binding partners present in the cell. Here we have suggested a simple sorting mechanism that promotes rapid identification of the right ligand by first quickly selecting all potential ligands that fall within the right category (here carrying the right charge), and then attempting to bind. It was found that in the presence of "wrong" ligands, binding is promoted most with a pre-binding site that has a low rather than a high affinity, in order to avoid creating kinetic traps with wrong ligands. This model is yet to be supported by experimental evidence and further sorting mechanisms, e.g. by size, shape, hydrophobicity etc. might exist.

Acknowledgments

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

References

Mazen Ahmad, Wei Gu, and Volkhard Helms. Mechanism of fast peptide recognition by sh3 domains. *Angew Chem Int Ed Engl*, 47(40):7626–30, Jan 2008.

G Altobelli and S Subramaniam. Kinetics of association of anti-lysozyme monoclonal antibody d44.1 and hen-egg lysozyme. *Biophys J*, 79(6):2954–65, Dec 2000.

Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.*, 53(1):291–318, 2002. ISSN 0066-426X.

Gregory R. Bowman, Kyle A. Beauchamp, George Boxer, and Vijay S. Pande. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.*, 131(12):124101+, September 2009. ISSN 1089-7690.

Martin Brune, Jackie L. Hunter, Steven A. Howell, Stephen R. Martin, Theodore L. Hazlett, John E. T. Corrie, and Martin R. Webb. Mechanism of inorganic phosphate interaction with phosphate binding protein from escherichia coli. *Biochemistry*, 37(29):10370–10380, July 1998.

Ignasi Buch, Toni Giorgino, and Gianni De Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 108(25):10184–9, Jun 2011.

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

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

John D. Chodera and Frank Noé. Probability distributions of molecular observables computed from markov models. ii: Uncertainties in observables and their time-evolution. *J. Chem. Phys.*, 133:105102, 2010.

B. de Groot, X. Daura, A. Mark, and H. Grubmüller. Essential Dynamics of Reversible Peptide Folding: Memory-free Conformational Dynamics Governed by Internal Hydrogen Bonds. *J. Mol. Bio.*, 301:299–313, 2001.

Daniele Dell'Orco. Fast predictions of thermodynamics and kinetics of protein-protein recognition from structures: from molecular design to systems biology. *Mol Biosyst*, 5(4):323–34, Apr 2009.

Ken A. Dill. Polymer principles and protein folding. *Protein Science*, 8(6):1166–1180, June 1999. ISSN 0961-8368.

Rose Du, Vijay S. Pande, Alexander Yu, Toyoichi Tanaka, and Eugene S. Shakhnovich. On the transition coordinate for protein folding. *The Journal of Chemical Physics*, 108(1):334–350, 1998.

L Ehrlich, C Münkel, G Chirico, and J Langowski. A brownian dynamics model for the chromatin fiber. *Comput Appl Biosci*, 13(3):271–9, Jun 1997.

Adrian H. Elcock. Atomic-level observation of macromolecular crowding effects: Escape of a protein from the groel cage. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2340–2344, March 2003.

Adrian H. Elcock, Razif R. Gabdoulline, Rebecca C. Wade, and J. Andrew McCammon. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J. Mol. Biol.*, 291(1):149–162, August 1999. ISSN 0022-2836.

Radek Erban and S. Jonathan Chapman. Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.*, 6(4):046001, December 2009. ISSN 1478-3975.

D Ermak and J McCammon. Brownian dynamics with hydrodynamic interactions. *J Chem Phys*, 69(4):1352–1360, Jan 1978.

R R Gabdoulline and R C Wade. Simulation of the diffusional association of barnase and barstar. *Biophys J*, 72(5):1917–29, May 1997a.

R R Gabdoulline and R C Wade. On the protein-protein diffusional encounter complex. *J Mol Recognit*, 12(4):226–34, Jan 1999.

R R Gabdoulline and R C Wade. Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. *J Mol Biol*, 306(5):1139–55, Mar 2001a.

Razif R. Gabdoulline and Rebecca C. Wade. Effective charges for macromolecules in solvent. *Journal of Physical Chemistry*, 100(9):3868–3878, February 1996.

Razif R. Gabdoulline and Rebecca C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophys. J.*, 72(5):1917–1929, May 1997b.

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

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

Razif R Gabdoulline and Rebecca C Wade. Biomolecular diffusional association. *Curr Opin Struct Biol*, 12(2):204–13, Apr 2002.

Tihamér Geyer and Uwe Winter. An o(n2) approximation for hydrodynamic interactions in brownian dynamics simulations. *J Chem Phys*, 130(11):114905, Mar 2009.

Esmael J Haddadian and Elizabeth L Gross. A brownian dynamics study of the interactions of the luminal domains of the cytochrome b6f complex with plastocyanin and cytochrome c6: the effects of the rieske fes protein on the interactions. *Biophys J*, 91(7):2589–600, Oct 2006.

Martin Held, Philipp Metzner, Jan-Hendrik Prinz, and Frank Noé. Mechanisms of protein-ligand association and its modulation by protein mutations. *Biophys J*, 100(3):701–710, Feb 2011.

Nina S. Hinrichs and Vijay S. Pande. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.*, 126: 244101, 2007.

Hung-Chung Huang and James M. Briggs. The association between a negatively charged ligand and the electronegative binding pocket of its receptor. *Biopolymers*, 63(4):247–260, 2002. ISSN 1097-0282.

Isaac A. Hubner, Eric J. Deeds, and Eugene I. Shakhnovich. Understanding ensemble protein folding at atomic detail. *Proc. Natl. Acad. Sci. USA*, 103(47):17747–17752, November 2006.

Mary E. Karpen, Douglas J. Tobias, and Charles L. Brooks. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry*, 32(2):412–420, January 1993.

W Kauzmann. Some factors in the interpretation of protein denaturation. *Adv Protein Chem*, 14:1–63, Jan 1959.

Bettina Keller, Jan-Hendrik Prinz, and Frank Noé. Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics. *Chem. Phys. (in revision)*, 2011.

Oliver F. Lange, Nils-Alexander Lakomek, Christophe Fares, Gunnar F. Schroder, Korvin F. A. Walter, Stefan Becker, Jens Meiler, Helmut Grubmuller, Christian Griesinger, and Bert L. de Groot. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. *Science*, 320(5882):1471–1475, June 2008.

Juan Latorre, Philipp Metzner, Carsten Hartmann, and Christof Schütte. A structure-preserving numerical discretization of reversible diffusions. *submitted to Comm. Math. Sci.*, 2010.

P. S. Ledvina, A. L. Tsai, Z. Wang, E. Koehl, and F. A. Quiocho. 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(12):2550–2559, December 1998. ISSN 0961-8368.

SR McGuffee and Adrian Elcock. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol*, 6(3):e1000694, 2010.

Robert H Meltzer, Errol Thompson, Kizhake V Soman, Xing-Zhi Song, Jerry O Ebalunode, Theodore G Wensel, James M Briggs, and Steen E Pedersen. Electrostatic steering at acetylcholine binding sites. *Biophys J*, 91(4):1302–14, Aug 2006.

Philipp Metzner, Christof Schütte, and Eric vanden Eijnden. Transition path theory for markov jump processes. *Multiscale Model. Simul.*, 7(3):1192–1219, 2009.

Stefanie Muff and Amedeo Caflisch. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a -sheet miniprotein. *Proteins*, 70:1185–1195, 2007.

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

Frank Noé and Stefan Fischer. Transition networks for modeling the kinetics of conformational transitions in macromolecules. *Curr. Opin. Struc. Biol.*, 18:154–162, 2008.

Frank Noé, Illia Horenko, Christof Schütte, and Jeremy C. Smith. Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.*, 126:155102, 2007.

Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the full ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, 106:19011–19016, 2009.

S H Northrup and H P Erickson. Kinetics of protein-protein association explained by brownian dynamics computer simulation. *Proc Natl Acad Sci USA*, 89(8):3338–42, Apr 1992.

Scott H. Northrup, Stuart A. Allison, and J. Andrew McCammon. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.*, 80(4):1517–1524, 1984.

José N. Onuchic and Peter G. Wolynes. Theory of Protein Folding. *Curr. Opin. Struc. Biol.*, 14:70–75, 2004.

Georgi V Pachov, Razif R Gabdoulline, and Rebecca C Wade. On the structure and dynamics of the complex of the nucleosome and the linker histone. *Nucleic Acids Res*, 39(12):5255–63, Jul 2011.

Albert C. Pan and Benoît Roux. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.*, 129(6):064107+, August 2008. ISSN 1089-7690.

Garegin A Papoian and Peter G Wolynes. The physics and bioinformatics of binding and folding-an energy landscape perspective. *Biopolymers*, 68(3):333–49, Mar 2003.

L Pauling, D Pressman, and D H Campbell. An experimental test of the framework theory of antigen-antibody precipitation. *Science*, 98(2542):263–4, Sep 1943.

Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: generation and validation. *J Chem Phys*, 134(17):174105, May 2011.

Jan-Hendrik Prinz, Bettina Keller, and Frank Noé. Probing molecular kinetics with markov models: Metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.*, 2011 (in revision).

Z Radic. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. distinctions between active center ligands and fasciculin. *Journal of Biological Chemistry*, 272(37):23265–23277, Sep 1997.

F. Rao and A. Caflisch. The Protein Folding Network. *J. Mol. Bio.*, 342:299–306, 2004.

D R Ripoll, C H Faerman, P H Axelsen, I Silman, and J L Sussman. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc Natl Acad Sci USA*, 90(11):5128–32, Jun 1993.

Jakob Schluttig, Denitsa Alamanova, Volkhard Helms, and Ulrich S. Schwarz. Dynamics of protein-protein encounter: A langevin equation approach with reaction patches. *J. Chem. Phys.*, 129(15):155106, 2008. ISSN 00219606.

G Schreiber, G Haran, and Huan-Xiang Zhou. Fundamental aspects of protein-protein association kinetics. *Chemical reviews*, 109(3):839–60, Mar 2009.

V. Schultheis, T. Hirschberger, H. Carstens, and P. Tavan. Extracting Markov Models of Peptide Conformational Dynamics from Simulation Data. *J. Chem. Theory Comp.*, 1:515–526, 2005.

C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A Direct Approach to Conformational Dynamics based on Hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.

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

N. Singhal, C. Snow, and V. S. Pande. Path sampling to build better roadmaps: predicting the folding rate and mechanism of a Trp Zipper beta hairpin. *J. Chem. Phys.*, 121:415–425, 2004.

Yuhua Song, Yongjie Zhang, Tongye Shen, Chandrajit L. Bajaj, Andrew A. McCammon, and Nathan A. Baker. Finite element solution of the steady-state smoluchowski equation for rate constant calculations. *Biophys. J.*, 86(4):2017–2029, April 2004. ISSN 00063495.

A. Spaar, C. Dammer, R. R. Gabdoulline, R. C. Wade, and V. Helms. Diffusional encounter of barnase and barstar. *Biophys J*, 90:1913–1924, 2006a.

A. Spaar, D. Flock, and V. Helms. Association of cytochrome c with membrane-bound cytochrome c oxidase proceeds parallel to the membrane rather than in bulk solution. *Biophys J*, 96:1721–1732, 2009.

Alexander Spaar, Christian Dammer, Razif R. Gabdoulline, Rebecca C. Wade, and Volkhard Helms. Diffusional encounter of barnase and barstar. *Biophys. J.*, 90(6):1913–1924, March 2006b.

ME Stroppolo, M Falconi, AM Caccuri, and A Desideri. Superefficient enzymes. *Cellular and Molecular Life Sciences*, 58(10):1451–1460, 2001.

W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, and M. Eleftheriou. Describing protein folding kinetics by molecular dynamics simulations: 2. Example applications to alanine dipeptide and beta-hairpin peptide. *Journal of Physical Chemistry B*, 108:6582–6594, 2004.

D. Thirumalai and S. A. Woodson. Kinetics of Folding of Proteins and RNA. *Acc. Chem. Res.*, 29:433–439, 1996.

Peter J. Tummino and Robert A. Copeland. Residence time of receptor-ligand complexes and its effect on biological function. *Biochemistry*, 47(20):5481–5492, May 2008. ISSN 1520-4995.

E. Vanden-Eijnden. Transition Path Theory. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, pages 453–493. 2006.

M Vijayakumar, K Y Wong, G Schreiber, A R Fersht, A Szabo, and Huan-Xiang Zhou. Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. *J Mol Biol*, 278(5):1015–24, May 1998.

Vincent A. Voelz, Gregory R. Bowman, Kyle Beauchamp, and Vijay S. Pande. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9. *J. Am. Chem. Soc.*, 132(5):1526–1528, February 2010.

D. J. Wales. *Energy Landscapes*. Cambridge University Press, Cambridge, 2003.

M. Weber. Improved Perron cluster analysis. *ZIB Report*, 03-04, 2003.

H X Zhou. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys J*, 64(6):1711–26, Jun 1993.

Huan-Xiang Zhou. On the calculation of diffusive reaction rates using brownian dynamics simulations. *J Chem Phys*, 92(5):3092, 1990.

Figures

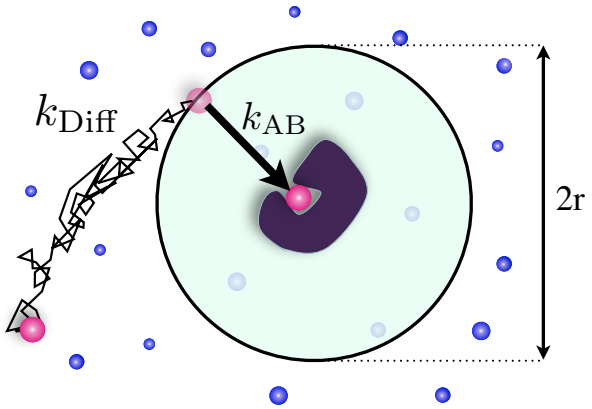

Figure 1: k_{Diff} and k_{AB} as contributors to the overall diffusional association rate k_{on} . k_{Diff} denotes the rate at which substrate particles diffuse into a sphere of radius r, it is given by the Smoluchowski rate $k_{\mathrm{Diff}} = 4\pi Dr$ and has the unit $M^{-1}s^{-1}$. k_{AB} denotes the transition rate of a single substrate molecule present at distance r to the binding site B, it is given by the $A \rightarrow B$ transition rate and has the unit s^{-1}.

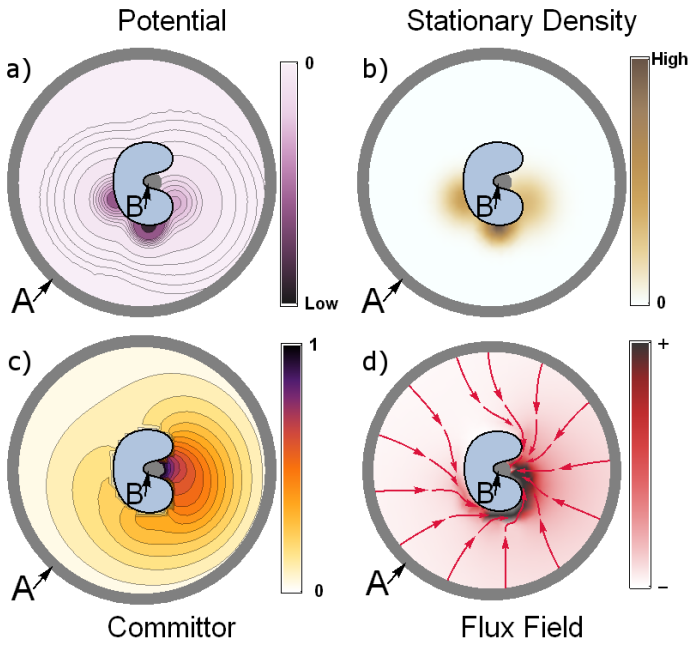

Figure 2: Illustration of TPT on a simple two-dimensional protein-ligand binding model. The dissociated state of the ligand A and the associated state B are shown. (a) potential energy landscape. (b) resulting stationary density, i.e., probability to find the ligand at a certain position (c) committor q, revealing a higher probability to reach the binding site for areas on top of the charged protein than for the uncharged protein. (d) reactive flux density and integrated flux lines calculated from the flux field resulting from the fluxes f_{ij}^{+}.

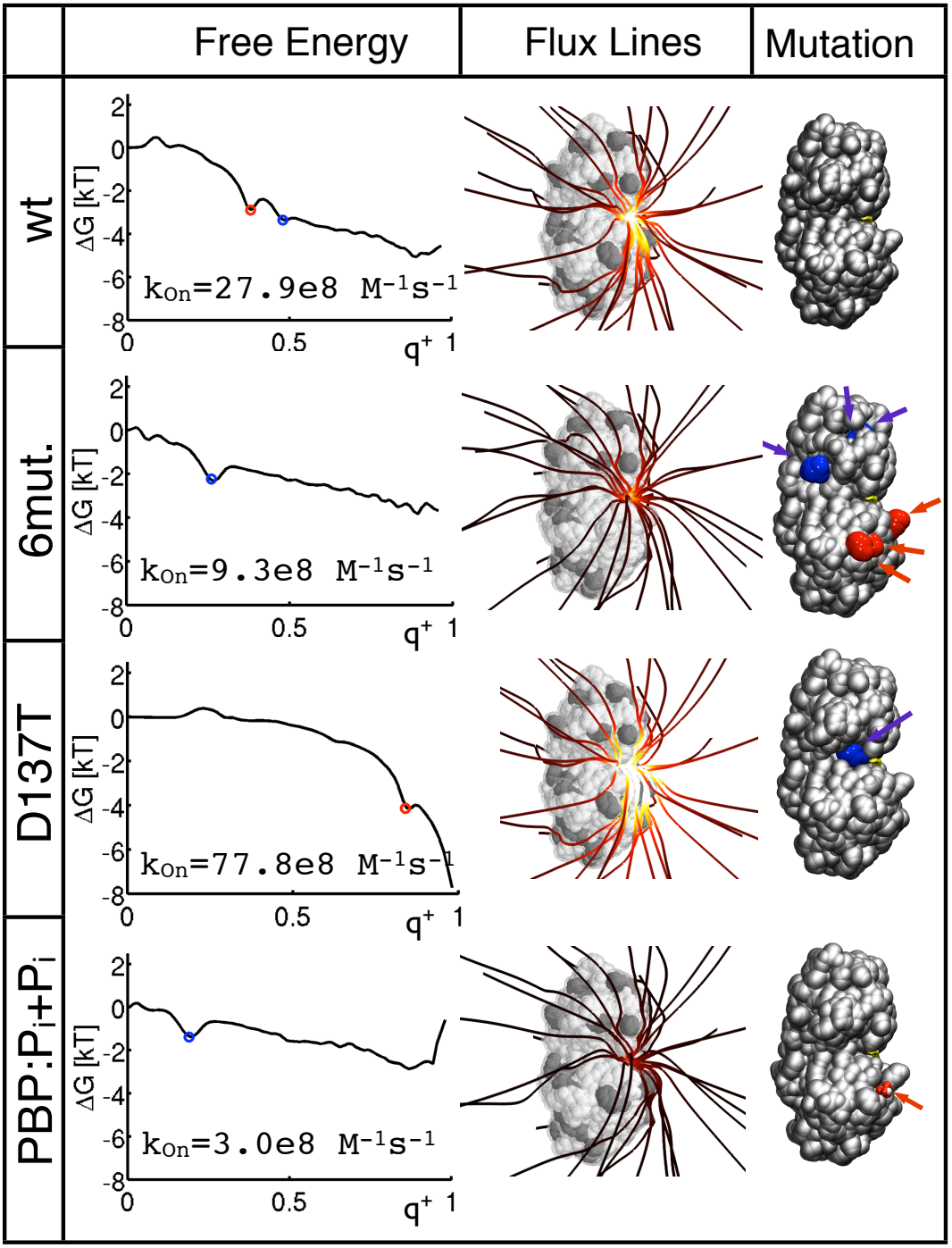

Figure 3: Free energy profiles and association pathways of P_i associating to the Phosphate Binding Protein. *First column* - Free energy profile of the ligand when it travels from the dissociated ($q = 0$) to the associated state ($q = 1$). *Second column* - Streamlines of the reactive flux of ligand association, which represent the ensemble of association pathways. A brighter streamline coloring corresponds to a higher local reactive flux. *Third column* - Here the positions of mutated residues are shown. Red/blue corresponds to negative/positive charged mutations relative to the wild-type structure.

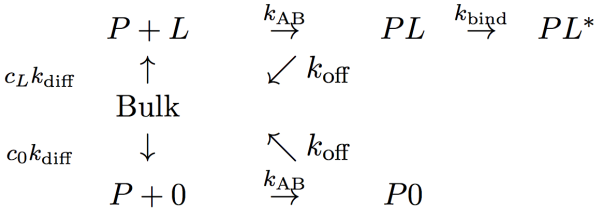

Figure 4: Reaction diagram showing a possible kinetic scenario of anion attraction to a pre-binding site. Starting from the bulk two possible reaction channels can be taken: In case of a ligand (L) the upper path is chosen, in case of a non-ligand (0) the lower one. Both species are attracted to the pre-binding site PL/P0 with their respective concentration dependent rate k_{on}, but only the ligand can move further to the bound state PL. If the concentration of non-ligand is increased with respect to the ligand concentration the lower channel will dominate the reaction system and the pre-binding site will be blocked by the non-ligand more often, which results in a lower yield of PL*.

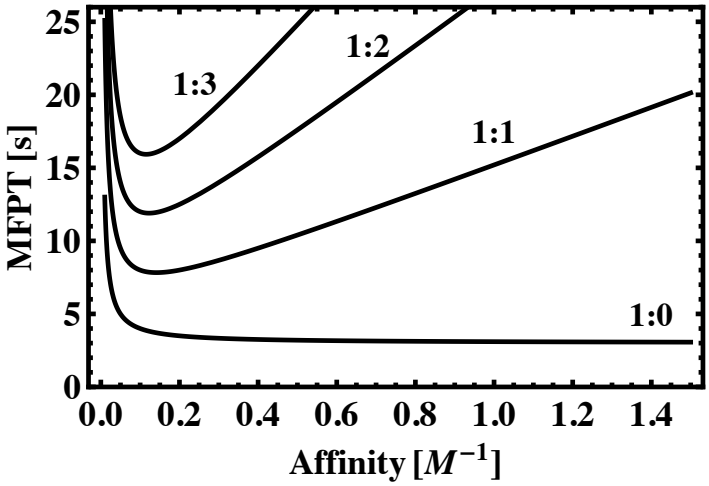

Figure 5: Mean binding time of a ligand in a mixture of "right" and "wrong" ligands at concentrations c_L and c_0 depending on the relative concentration $c_L : c_0$ and the pre-binding affinity $K_a = k_{on}/k_{off}$. Rates were set to $(c_0 + c_L)k_{\text{diff}} = k_{AB} = k_{bind} = k_u = 1s^{-1}$ and $k_{off} = k_{on}/K_a$.