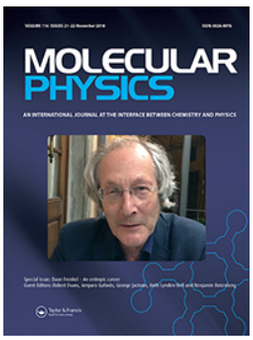


Taylor & Francis
Taylor & Francis Group

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: http://www.tandfonline.com/loi/tmph20

Generalised expressions for the association and dissociation rate constants of molecules with multiple binding sites

Adithya Vijaykumar, Pieter Rein ten Wolde & Peter G. Bolhuis

To cite this article: Adithya Vijaykumar, Pieter Rein ten Wolde & Peter G. Bolhuis (2018) Generalised expressions for the association and dissociation rate constants of molecules with multiple binding sites, Molecular Physics, 116:21-22, 3042-3054, DOI: 10.1080/00268976.2018.1473653

To link to this article: https://doi.org/10.1080/00268976.2018.1473653

Published online: 25 May 2018.

MOLECULAR PHYSICS
2018, VOL. 116, NOS. 21–22, 3042–3054
https://doi.org/10.1080/00268976.2018.1473653

Taylor & Francis
Taylor & Francis Group

FRENKEL SPECIAL ISSUE

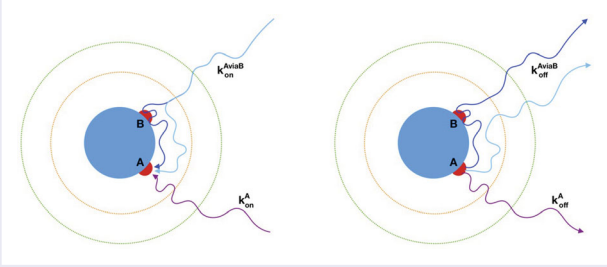
Generalised expressions for the association and dissociation rate constants of molecules with multiple binding sites

Adithya Vijaykumar[a,b], Pieter Rein ten Wolde[a] and Peter G. Bolhuis[b]

[a]NWO Institute AMOLF, Amsterdam, Netherlands; [b]van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

Many proteins exhibit multiple binding patches. A patch may harbour a key chemical modification site, but may also simply act as a trap for the binding to another site. Here we consider the scenario in which one molecule (enzyme) binds another molecule (substrate) which contains two sites. We present microscopic expressions for the rate at which the enzyme binds to a particular site on the substrate, both for the scenario in which the enzyme directly binds the site without first visiting the other site, and for the case in which it may visit the other site an arbitrary number of times before binding to the site of interest. We also present the expressions for the corresponding dissociation reactions. These expressions can be used to compute in a single rare-event simulation of the dissociation pathway not only both the intrinsic and effective dissociation rate constants but also both association rate constants.



ARTICLE HISTORY
Received 12 March 2018
Accepted 20 April 2018

KEYWORDS
Biochemical networks; binding kinetics; transition path theory; path sampling; forward flux sampling

1. Introduction

Arguably the most elementary event inside the living cell is a chemical reaction, be it a metabolic reaction, a chemical modification of a signalling protein, the binding of a gene regulatory protein to the DNA, or the hydrolysis of a fuel molecule that allows a motor protein to generate forces or move along a filament. Chemical reactions often involve the association and dissociation of two molecules. Some processes are controlled by the equilibrium constant of the association–dissociation reaction – the ratio of the association rate constant over the dissociation rate constant. For example, the average level of gene expression is, to a good approximation, determined by the binding probability of the gene regulatory protein to the promoter [1]. However, the living cell is a highly non-equilibrium system, and many, if not most, processes are governed by the absolute values of the rate constants. In protein signalling, the pathway for protein demodification is not the microscopic reverse of protein modification, setting up a cycle of modification and demodification. Not only the balance between modification and demodification, which sets the modification level (the output signal), but also the flux around the cycle, which determines the fuel consumption, is governed by the absolute rates of protein modification and demodification [2]. Many processes, such as cell division and cell differentiation during development, are directed by spatial protein patterns, and a key parameter in the formation of these patterns is the diffusion constant over the reaction rate. The slow binding of transcription factors to the DNA can be a major source of gene expression [3] and limit the speed of the response to a change in the environment [4]. Last but not least, recent studies indicate that the discrimination by the immune response does

CONTACT Peter G. Bolhuis p.g.bolhuis@uva.nl van't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157, 1090 GD Amsterdam, Netherlands

not rely on the binding probability of the antigen to the T-cell receptor, but rather on their dissociation rate [5].

The binding between two molecules that interact in an isotropic fashion or via a single site only has been studied in great detail. Theories have been developed that yield expressions for the overall association and dissociation rate constants in terms of the cross sections, interaction potentials, diffusion constants and intrinsic rate constants [6,7]. Methods have been developed that make it possible to compute rate constants in simulations of more complicated systems [8,9]. Recently, we have presented an efficient scheme that makes it possible to compute in a single rare-event simulation of the dissociation pathway, using, e.g. transition interface sampling (TIS) [10,11] or forward flux sampling (FFS) [12,13], the effective and intrinsic rate constants for both dissociation and association [14,15]. The scheme is based on the observation that the escape probability $P_{\rm esc} = k_D/(k_a + k_D)$, being a splitting probability, not only depends on the diffusion-limited arrival rate $k_{\rm D}$, which is known analytically, but also on the intrinsic association rate $k_{\rm a}$ [14]. Hence, by computing the escape probability $P_{\rm esc}$ during a TIS/FFS simulation of the dissociation process, we obtain information on the intrinsic association rate $k_{\rm a}$, and hence, also on the effective association rate constant.

Yet, many proteins exhibit not one binding site but many, often because these proteins can be chemically modified at multiple sites. Indeed, multi-site protein modification is a very common motif in biology. Perhaps the most studied and best characterised example is the mitogen-activated protein kinase system [16], but other well-known examples are the circadian clocks of cyanobacteria [17], fungi [18], higher order organisms [19], the Nuclear Factor of Activated T-cells system [20] and the CaM kinase II system [21]. It is well known that the response of these systems strongly depends on whether the different sites are modified in a processive fashion, where the enzyme and substrate remain bound in between the modification steps, or via a distributive scheme, where the enzyme and substrate dissociate from each other after each modification step. A distributive scheme can give rise to an ultrasensitive response [22,23], to a threshold response [24] or to bistability [21,25]. However, a scheme that is distributive can, due to rapid enzyme–substrate rebindings, effectively become processive, leading to the loss of ultra-sensitivity and bistability [16,26,27]. Hence, the dynamics of these systems does not only depend on the rates at which the enzyme and substrate associate with each other coming from the bulk and dissociate from one another back into the bulk, but also on the rate at which the enzyme can rebind the substrate, i.e. hop to another site on the same substrate molecule [16].

Recently, we have demonstrated how the scheme of [14] can be generalised to particles with multiple patches [28]. This scheme makes it possible to compute not only the rate at which two molecules (un)bind from (into) the bulk, but also the rate at which one rebinds the other, i.e. the rate at which a molecule dissociates from one site on the other molecule and then hops to and binds to another site. Importantly, however, in the scheme presented in [28], the rate of binding to a particular site was taken to be the rate at which the enzyme, coming from the bulk, binds that site *without* first binding the other patch as an intermediate step. This seems to be the most natural quantity for the multi-site protein modification systems listed above.

Yet, in many cases one site is the site of interest, while the other merely acts as a trap. This is a common scenario in protein–protein association, but also in the binding of gene regulatory proteins to the DNA. Here, the quantity of interest is the rate of binding to a particular site, while allowing for the intermediate binding to the other site. The quantity of interest may also be dictated by the experimental setup; using, for example, FRET, one may measure the rate at which a molecule binds to a particular site on another molecule, irrespective of how many times it has visited other sites before binding to the site of interest. In this paper, we show how the scheme of [28] can be generalised to enable not only the computation of the rate constant of direct association and dissociation, but also that which allows the intermediate binding to other sites.

While we phrased the problem in terms of biomolecular association, we note that this approach is generally applicable to all kinds of dissociation and association processes, including chemical reactions, structure formation of colloidal particles or nanoparticles, and surfactant assembly.

This paper is organised as follows. In the next section, we derive along similar lines as in [28] generalised expressions for the case of intermediate binding. In Section 3, we apply the technique to the system also studied in [28]. We discuss the results and give concluding remarks in the final section.

2. Theory

2.1. The path ensembles

In previous work [28], we considered binding and unbinding of an enzyme with a substrate with two competing patches A and B, where both patches were treated on equal footing, and were considered a target state. We showed how from a single FFS or TIS simulation of a dissociation reaction from patch A, we can compute

both the intrinsic and effective dissociation rate constants, denoted k_d^{A} and $k_{\mathrm{off}}^{\mathrm{A}}$, respectively, the intrinsic and effective hopping rates between A and B, as well as the intrinsic and effective association rate constants to patch A, denoted k_a^{A} and $k_{\mathrm{on}}^{\mathrm{A}}$, respectively. In addition, we gave expressions for $k_a^{\mathrm{A}\vee\mathrm{B}}$ and $k_{\mathrm{on}}^{\mathrm{A}\vee\mathrm{B}}$, respectively, the intrinsic and effective association rate constant, for the case of binding to either of the two (identical) patches.

Here we consider the case of a substrate with two patches in which one of the patches (B) can be considered an intermediate state, and binding to A can be described by two-state kinetics. This is the case when the dissociation constant K_{eq}^A for binding to A is smaller than that for binding to B, $K_{\mathrm{eq}}^A \ll K_{\mathrm{eq}}^B$, i.e. the substrate binds much stronger to A than to B. Still B can be considered a binding site by itself, i.e. the residence time of B is long on the molecular timescale. Association and dissociation to patch A can thus involve one or more temporary bindings to B, and can be described with effective dissociation and association rate constants that account for intermediate binding to B. In this section, we derive microscopic expressions for the intrinsic and effective association, dissociation and hopping rates for this situation. We will first discuss the path ensembles that correspond to each of the rate constants.

The intrinsic dissociation rate constants for the enzyme initially bound at patch A, $k_{\mathrm{d}}^{\mathrm{AviaB}}(\sigma)$, involves counting the trajectories that start at A and reach σ either directly without reaching the patch B, or first visit B (possibly multiple times) and then go to σ (see Figure 1 for a graphical illustration). Similarly, the effective dissociation rate constant for the enzyme starting at patch A, $k_{\mathrm{off}}^{\mathrm{AviaB}}$, takes into account all trajectories starting at A and going to infinity, with or without visiting patch B.

The effective association constant of an enzyme to bind to patch A, $k_{\mathrm{on}}^{\mathrm{AviaB}}$, corresponds to trajectories that start at infinity and terminate at patch A with or without visiting B, while the intrinsic association rate constant $k_{\mathrm{a}}^{\mathrm{AviaB}}(\sigma)$ corresponds to trajectories that start at the σ-interface and terminate at patch A with or without first visiting B.

In [28], we also distinguish the intrinsic and effective rate constant of hopping from patch A to B, $k_{\mathrm{hop}}^{AB}(\sigma)$, $k_{\mathrm{effHop}}^{AB}(\sigma)$, which corresponds to trajectories that start in patch A and end in patch B, with or without first visiting the σ-interface. We do not consider these rates explicitly here, although we do take the hopping process into account in the association and dissociation rates.

We also emphasise that while the intrinsic rate constants depend on the definition of σ, the effective rate constants do not.

All the rate constants (intrinsic/effective association and dissociation) can be obtained from one single FFS (or TIS) simulation of a dissociation reaction. To see how, it is instructive to imagine that starting from the bound state A, we have generated N configurations at interface σ. These configurations

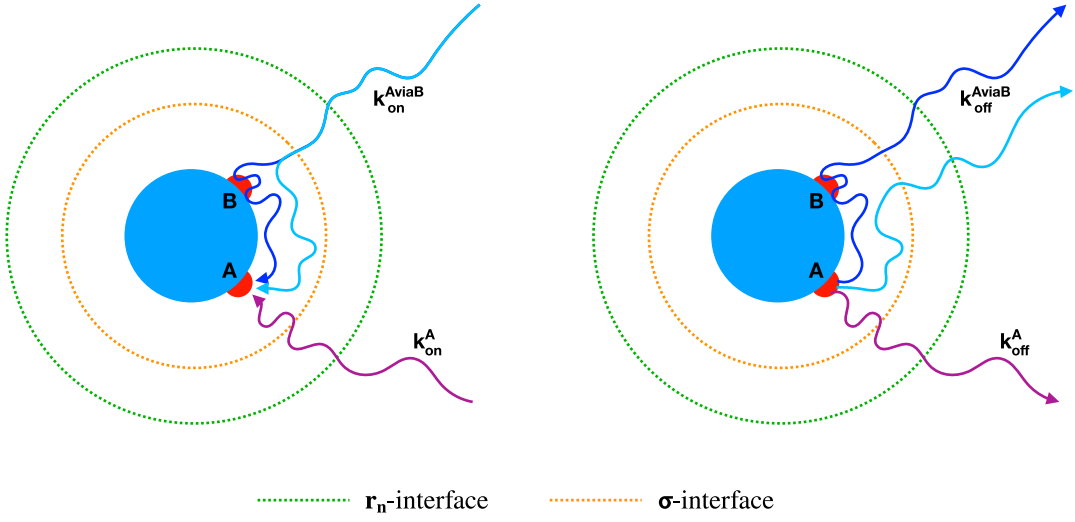


Figure 1. The path ensembles contributing to the effective association rate constants (left) and the effective dissociation rate constants (right). The substrate has two patches A and B and when the particle diffuses to infinity, it is in the unbound state. All rates are calculated with respect to patch A, i.e. the rate constant of binding to patch A or the rate constant of unbinding from patch A. Thus, $k_{\mathrm{on}}^{\mathrm{AviaB}}$ is the association rate constant to patch A, and includes direct paths as well as paths via B. k_{on}^{A} is the association rate constant for a single patch A and (obviously) only includes direct paths. The analogously defined effective off rates are illustrated in the panel on the right. For clarity we did not show the rates to and from either of the two patches $k_{\mathrm{on}}^{A\vee B}$ and $k_{\mathrm{off}}^{\mathrm{A}\vee\mathrm{B}}$. σ is the interface where the intrinsic rate constants (not shown) are measured and r_{n} is an interface beyond σ.

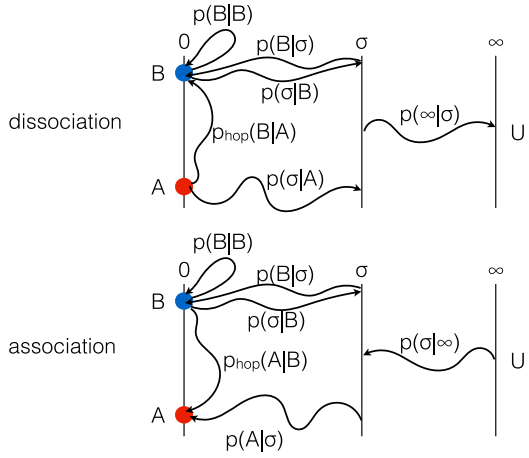


Figure 2. Overview of the relevant conditional probabilities in dissociation (top) and association (bottom) process for substrates with multiple binding sites A and B. Shown are the relevant basic path types and corresponding conditional probabilities required to construct the effective and intrinsic rate constants. For simplicity the interfaces are shown in a one-dimensional fashion. The r_n interface is not shown.

are thus distributed over the σ surface according to the distribution as obtained from an FFS dissociation simulation starting from patch A. Of these N trajectories at interface σ, $N_{\sigma\to B}$ progress (on average) to B, $N_{\sigma\to A}$ return to A and $N_{\sigma\to\infty}$ dissociate, i.e. escape to infinity: $N = N_{\sigma\to \mathrm{B}} + N_{\sigma\to \mathrm{A}} + N_{\sigma\to\infty}$. In the limit of $N\to\infty$, we can define the probabilities $P(B \mid \sigma) = N_{\sigma\to B}/N$, $P(A \mid \sigma) = N_{\sigma\to A}/N$, $P(\infty \mid \sigma) = N_{\sigma\to\infty}/N$. For a finite number of sampled trajectories, the trajectory fractions become approximations of these probabilities.

FFS also allows constructing N trajectories starting from A – as described by the first interface r_0^A – and ending in B, returning to A or reaching σ. Of the N trajectories emanating from the FFS interface r_0^A delimiting state A $N_{A\to B}$ progress (on average) to B, $N_{A\to A}$ return to A and $N_{A\to\sigma}$ dissociate to reach σ: $N = N_{\mathrm{A}\to\mathrm{B}} + N_{\mathrm{A}\to\mathrm{A}} + N_{\mathrm{A}\to\sigma}$. In the limit of $N\to\infty$, we can define the probabilities $P(B \mid A) = N_{A\to B}/N$, $P(A \mid A) = N_{A\to A}/N$, $P(\infty \mid A) = N_{A\to\sigma}/N$. In addition, an FFS run starting from the r_0^B interface at state B will give rise to $N_{B\to\sigma}$ trajectories to the σ-surface, $N_{B\to B}$ trajectories returning to B and $N_{B\to A}$ trajectories hopping to A. Again, in the limit of $N\to\infty$, we can define the probabilities $P(\sigma \mid B) = N_{B\to\sigma}/N, P(B \mid B) = N_{B\to B}/N$, $P(A \mid B) = N_{B\to A}/N$. Note that we replaced the interfaces r_0^A and r_0^B, by the shorthand A and B, respectively. The above conditional probabilities are shown graphically in Figure 2.The different conditional probabilities for trajectories starting at the σ-surface are also indicated.

2.2. The effective dissociation rate constant

We first focus on the dissociation process. The intrinsic dissociation rate constant at the σ-interface is given by

$$k_{\mathrm{d}}^{\mathrm{AviaB}}(\sigma) = \Phi_0 P_A(\sigma \mid r_0^A), \tag{1}$$

where, as before, Φ_0 is the flux across the first FFS interface r_0^A, and $P_A(\sigma \mid r_0^A)$ is the probability that an enzyme starting at this first interface r_0 of state A reaches the σ-interface before returning to A, and can include one or more visits to B. The effective dissociation constant can then be expressed as,

$$\begin{aligned} k_{\mathrm{off}}^{\mathrm{AviaB}} &= \Phi_0 P_A(\sigma \mid r_0^A) P^{\mathrm{eff}}(\infty \mid \sigma) \\ &= k_{\mathrm{d}}^{\mathrm{AviaB}}(\sigma) P^{\mathrm{eff}}(\infty \mid \sigma), \end{aligned} \tag{2}$$

where $P^{\mathrm{eff}}(\infty \mid \sigma)$ is the probability that a trajectory that has reached σ escapes to infinity, i.e. dissociates, before it (re)associates with patch A. Thus, this probability also includes trajectories with one or more visits to B.

The probabilities of interest $P_A(\sigma \mid r_0^A)$ and $P^{\mathrm{eff}}(\infty \mid \sigma)$ cannot be computed directly, as both include many paths with B as intermediate. This means that the paths via state B are (temporarily) trapped, but then can either hop back to A, or dissociate and rebind to B, or can escape to σ. We do not want to simulate those paths explicitly, as B is potentially a strongly bound state, and trajectory times can become extremely long. However, we can decompose the relevant paths into different categories, each with their own probability as indicated in Figure 2: (1) direct paths from A to σ, with probability $P(\sigma \mid A)$, (2) indirect paths from B to σ with probability $P(\sigma \mid B)$, (3) rebinding paths from B to B with probability $P(B \mid B)$, (4) paths that leave σ and reach B with probability $P(B \mid \sigma)$, (5) hopping paths from A to B with probability $P_{\mathrm{hop}}(B \mid A)$. Note that we have again replaced the interfaces r_0^A and r_0^B, by the shorthand A and B, respectively. Thus, the probability $P_{\mathrm{A}}(\sigma \mid r_0^A)$ in Equation (2) includes both direct paths and indirect paths via B

$$P_{\mathrm{A}}(\sigma \mid r_0^A) = P(\sigma \mid A) + P^{\mathrm{AB}\sigma}(\sigma \mid A), \tag{3}$$

where $P^{\mathrm{AB}\sigma}(\sigma \mid A)$, the probability for paths that visit B, rebind an arbitrary number of times to B, before going back to A. This probability is given by the geometrical series

$$\begin{aligned} P^{\mathrm{AB}\sigma}(\sigma \mid A) = {} & P(\sigma \mid B) P_{\mathrm{hop}}(B \mid A) \\ & + P(\sigma \mid B) P(B \mid B) P_{\mathrm{hop}}(B \mid A) \\ & + P(\sigma \mid B) P(B \mid B)^2 P_{\mathrm{hop}}(B \mid A) + \cdots \end{aligned} \tag{4}$$

which can be re-summed to give

$$P^{\mathrm{AB}}\sigma(\sigma \mid A) = P(\sigma \mid B)P_{\mathrm{hop}}(B \mid A)\frac{1}{1-P(B \mid B)}. \quad (5)$$

Since we know that paths from B can only go to A, B or σ, it holds that $1 - P(B \mid B) = P(\sigma \mid B) + P_{\mathrm{hop}}(A \mid B)$ giving

$$\begin{aligned} P^{\mathrm{AB}}\sigma(\sigma \mid A) &= P_{\mathrm{hop}}(B \mid A)\frac{P(\sigma \mid B)}{P(\sigma \mid B) + P_{\mathrm{hop}}(A \mid B)} \\ &\equiv P_{\mathrm{hop}}(B \mid A)P^{\mathrm{norm}}(\sigma \mid B), \end{aligned} \quad (6)$$

where in the last equality we defined the normalised probability $P^{\mathrm{norm}}(\sigma \mid B)$ to reach σ from B in an arbitrary number of rebinding steps. Note that this is in fact a type of splitting probability or committor. Thus, the probability $P_A(\sigma \mid r_0^A)$ in Equation (2) is

$$P_{\mathrm{A}}(\sigma \mid r_0^A) = P(\sigma \mid A) + P^{\mathrm{hop}}(B \mid A)P^{\mathrm{norm}}(\sigma \mid B). \quad (7)$$

The probability $P^{\mathrm{eff}}(\infty \mid \sigma)$ in Equation (2) also includes direct paths and indirect paths with one or multiple visits to B. Similar to $P^{\mathrm{AB}}\sigma(\sigma \mid A)$, we can write for the indirect paths going from σ to σ, visiting B an arbitrary number of times without visiting A,

$$\begin{aligned} P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma) &= P(B \mid \sigma)\frac{P(\sigma \mid B)}{P(\sigma \mid B) + P_{\mathrm{hop}}(A \mid B)} \\ &\equiv P(B \mid \sigma)P^{\mathrm{norm}}(\sigma \mid B). \end{aligned} \quad (8)$$

The probability $P^{\mathrm{eff}}(\infty \mid \sigma)$ is governed by paths that shuttle back and forth between σ and B an arbitrary number of times before they escape to infinity. Again this leads to a geometrical series

$$\begin{aligned} P^{\mathrm{eff}}(\infty \mid \sigma) = P(\infty \mid \sigma) &+ P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma)P(\infty \mid \sigma) \\ &+ P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma)^2 P(\infty \mid \sigma)\cdots \end{aligned} \quad (9)$$

which again can be re-summed to give

$$P^{\mathrm{eff}}(\infty \mid \sigma) = \frac{P(\infty \mid \sigma)}{1 - P^{\sigma\mathrm{B}\sigma}} = \frac{P(\infty \mid \sigma)}{1 - P(B \mid \sigma)P^{\mathrm{norm}}(\sigma \mid B)}. \quad (10)$$

Combining Equations (2), (7) and (10), the effective dissociation rate from A is then

$$k_{\mathrm{off}}^{\mathrm{AviaB}} = \Phi_0 \frac{P(\sigma \mid A) + P^{\mathrm{hop}}(B \mid A)P^{\mathrm{norm}}(\sigma \mid B)}{1 - P(B \mid \sigma)P^{\mathrm{norm}}(\sigma \mid B)} P(\infty \mid \sigma). \quad (11)$$

2.3. Correction for using a finite r_n interface

Equation (11) gives the explicit expression for the effective rate constants, given the probabilities $P(B \mid \sigma)$, $P_{\mathrm{hop}}(B \mid A)$, $P(\sigma \mid B)$ and $P(\infty \mid \sigma)$. While $P_{\mathrm{hop}}(B \mid A)$, $P(\sigma \mid A)$ and $P(\sigma \mid B)$ are directly available in the FFS/TIS simulations from A and B, the $P(A \mid \sigma)$, $P(B \mid \sigma)$ and $P(\infty \mid \sigma)$ are not known directly, as they would require evaluating trajectories that go into the bulk to infinity. Following [28] we briefly summarise how $P(A \mid \sigma)$, $P(B \mid \sigma)$ and $P(\infty \mid \sigma)$ can be extracted from the simulations. As a start we write

$$\begin{aligned} P(B \mid \sigma) &= \frac{P(B \mid \sigma)}{P(A \mid \sigma) + P(B \mid \sigma)}(P(A \mid \sigma) + P(B \mid \sigma)) \\ &= \frac{P(B \mid \sigma)}{P(A \mid \sigma) + P(B \mid \sigma)}(1 - P(\infty \mid \sigma)) \\ &= \alpha(1 - P(\infty \mid \sigma)). \end{aligned} \quad (12)$$

where we have made use of the fact that $P(A \mid \sigma) + P(B \mid \sigma) + P(\infty \mid \sigma) = 1$, and we have introduced

$$\alpha \equiv \frac{P(B \mid \sigma)}{P(A \mid \sigma) + P(B \mid \sigma)} \quad (13)$$

as a 'splitting probability' for trajectories from σ arriving at B versus A. To compute $P(\infty \mid \sigma)$ in a brute-force manner, one would have to generate extremely long trajectories, because there is always a small but finite probability that an enzyme molecule which has diffused far away and deep into the bulk will return to the substrate molecule. To alleviate this problem, we put, following our earlier work [14], an extra interface at a position $r_{\mathrm{n}} > \sigma$. As we will show below, this extra interface makes it possible to efficiently compute $P(\infty \mid \sigma)$. Moreover, the probability $P(B \mid \sigma)$ for trajectories that move from σ to B is then given by the sum of the probability $P_{\mathrm{dir}}(B \mid \sigma)$ of trajectories that directly go from σ to B without first visiting r_{n}, and the probability $P_{r_{\mathrm{n}}}(B \mid \sigma)$ for those that first visit r_{n} and then proceed to B,

$$\begin{aligned} P(B \mid \sigma) &= P_{\mathrm{dir}}(B \mid \sigma) + P_{r_{\mathrm{n}}}(B \mid \sigma) \\ &= \alpha_{\mathrm{dir}}\,(P_{\mathrm{dir}}(A \mid \sigma) + P_{\mathrm{dir}}(B \mid \sigma)) \\ &\quad + \alpha_{r_{\mathrm{n}}}(P_{r_{\mathrm{n}}}(A \mid \sigma) + P_{r_{\mathrm{n}}}(B \mid \sigma)) \\ &= \alpha_{\mathrm{dir}}(1 - P(r_{\mathrm{n}} \mid \sigma)) \\ &\quad + \alpha_{r_{\mathrm{n}}}P(r_{\mathrm{n}} \mid \sigma)(1 - P(\infty \mid r_{\mathrm{n}})). \end{aligned} \quad (14)$$

Here, α_{dir} and $\alpha_{r_{\mathrm{n}}}$ are, respectively, the splitting probabilities of arriving in A versus B of those trajectories that proceed directly from σ to either A or B and those that arrive at A or B passing via r_{n}:

$$\alpha_{\mathrm{dir}} = \frac{P_{\mathrm{dir}}(B \mid \sigma)}{P_{\mathrm{dir}}(A \mid \sigma) + P_{\mathrm{dir}}(B \mid \sigma)} \quad (15)$$

$$\alpha_{r_{\mathrm{n}}} = \frac{P_{r_{\mathrm{n}}}(B \mid \sigma)}{P_{r_{\mathrm{n}}}(A \mid \sigma) + P_{r_{\mathrm{n}}}(B \mid \sigma)}. \quad (16)$$

Similarly for the trajectories starting at the σ-interface and reaching A, we can write

$$P(A \mid \sigma) = (1-\alpha_{\mathrm{dir}})(1-P(r_{\mathrm{n}} \mid \sigma)) + (1-\alpha_{r_{\mathrm{n}}})P(r_{\mathrm{n}} \mid \sigma)(1-P(\infty \mid r_{\mathrm{n}})). \quad (17)$$

Combining Equations (13), (14) and (17) yields

$$\alpha = \frac{\alpha_{\mathrm{dir}}(1-P(r_{\mathrm{n}} \mid \sigma)) + \alpha_{r_{\mathrm{n}}}P(r_{\mathrm{n}} \mid \sigma)(1-P(\infty \mid r_{\mathrm{n}}))}{1-P(\infty \mid \sigma)}. \quad (18)$$

Not only $P(r_{\mathrm{n}} \mid \sigma)$ and α_{dir}, but also $P(\infty \mid \sigma)$ can be obtained from the FFS simulations, as described below. Yet, to close the above equation and find $P(B \mid \sigma)$ (see Equation (12)), we need an expression for $\alpha_{r_{\mathrm{n}}}$. Since $P_{r_{\mathrm{n}}}(B \mid \sigma)$ is the product of the probability $P(r_{\mathrm{n}} \mid \sigma)$ of trajectories going from σ to r_{n} and the probability $P(B \mid r_{\mathrm{n}})$ of subsequently reaching B, the splitting probability $\alpha_{r_{\mathrm{n}}}$ in Equation (16) is also given by

$$\begin{aligned}\alpha_{r_{\mathrm{n}}} &= \frac{P_{r_{\mathrm{n}}}(B \mid \sigma)}{P_{r_{\mathrm{n}}}(A \mid \sigma) + P_{r_{\mathrm{n}}}(B \mid \sigma)} \\ &= \frac{P(r_{\mathrm{n}} \mid \sigma)P(B \mid r_{\mathrm{n}})}{P(r_{\mathrm{n}} \mid \sigma)P(A \mid r_{\mathrm{n}}) + P(r_{\mathrm{n}} \mid \sigma)P(B \mid r_{\mathrm{n}})} \\ &= \frac{P(B \mid r_{\mathrm{n}})}{P(A \mid r_{\mathrm{n}}) + P(B \mid r_{\mathrm{n}})}. \end{aligned} \quad (19)$$

We emphasise that up to this point, no assumption has been made. In particular, the expressions hold for any choice of the location σ, including one that is close to the bound state, which would lead to a non-uniform distribution of configurations at σ. With such a non-uniform distribution, $\alpha_{r_{\mathrm{n}}}$ is likely to be unequal to α_{dir}, which would make it impossible to close Equation (18). In contrast, if the distributions at the σ and the r_{n} interfaces are isotropic, then

$$\frac{P_{r_{\mathrm{n}}}(B \mid \sigma)}{P_{r_{\mathrm{n}}}(A \mid \sigma) + P_{r_{\mathrm{n}}}(B \mid \sigma)} \simeq \frac{P_{\mathrm{dir}}(B \mid \sigma)}{P_{\mathrm{dir}}(A \mid \sigma) + P_{\mathrm{dir}}(B \mid \sigma)}, \quad (20)$$

and thus

$$\alpha_{r_{\mathrm{n}}} = \alpha_{\mathrm{dir}}. \quad (21)$$

Inserting Equation (21) in Equation (18), we find

$$\alpha = \alpha_{r_{\mathrm{n}}} = \alpha_{\mathrm{dir}} \equiv \frac{P(B \mid \sigma)}{P(A \mid \sigma) + P(B \mid \sigma)}, \quad (22)$$

which turns Equation (12) to

$$P(B \mid \sigma) = \alpha_{\mathrm{dir}}(1-P(\infty \mid \sigma)). \quad (23)$$

Likewise it follows that

$$P(A \mid \sigma) = (1-\alpha_{\mathrm{dir}})(1-P(\infty \mid \sigma)). \quad (24)$$

Note that when A and B are identical patches, then $\alpha = \alpha_{\mathrm{dir}} = 0.5$. However, in general, this does not need to be the case. Indeed, in this work we explicitly assume that A is different from B.

As we assume $\alpha = \alpha_{\mathrm{dir}}$, from here onwards we will drop the subscript from α_{dir}. We stress, however, that computing α always amounts to computing α_{dir} as only that is directly available in the FFS/TIS simulations.

The only thing missing is an efficient computation of the escape probability $P(\infty \mid \sigma)$. In [28] we have shown that when introducing the r_n interface, this probability can be expressed as

$$P(\infty \mid \sigma) = \frac{P(r_{\mathrm{n}} \mid \sigma)(1-\Omega)}{1-\Omega P(r_{\mathrm{n}} \mid \sigma)} \quad (25)$$

or

$$1-P(\infty \mid \sigma) = \frac{1-P(r_{\mathrm{n}} \mid \sigma)}{1-P(r_{\mathrm{n}} \mid \sigma)\Omega}, \quad (26)$$

where $\Omega \equiv k_{\mathrm{D}}(\sigma)/k_{\mathrm{D}}(r_{\mathrm{n}})$ and $P(r_{\mathrm{n}} \mid \sigma)$ is the probability to reach r_n from σ, which follows from the FFS calculation. If the interfaces σ and r_{n} are beyond the cut-off of the potential, then the particles move by free diffusion. In this case, we can exploit the analytic expression for the diffusion-limited arrival rate constant $k_{\mathrm{D}}(\sigma) = 4\pi\sigma D$ to evaluate all the rate constants. Note that we have assumed here, as above, isotropic distributions at both r_n and σ [14].

Thus, the desired probabilities $P(A \mid \sigma)$ and $P(B \mid \sigma)$ follow from Equations (24) and (23)

$$P(B \mid \sigma) = \alpha\frac{1-P(r_{\mathrm{n}} \mid \sigma)}{1-P(r_{\mathrm{n}} \mid \sigma)\Omega} \quad (27)$$

$$P(A \mid \sigma) = (1-\alpha)\frac{1-P(r_{\mathrm{n}} \mid \sigma)}{1-P(r_{\mathrm{n}} \mid \sigma)\Omega}. \quad (28)$$

Defining $\beta = 1-P(\infty \mid \sigma)$ the final explicit expression for the effective off rate is

$$k_{\mathrm{off}}^{\mathrm{AviaB}} = \Phi\frac{P(\sigma \mid A) + P^{\mathrm{hop}}(B \mid A)P^{\mathrm{norm}}(\sigma \mid B)}{1-\alpha\beta P^{\mathrm{norm}}(\sigma \mid B)}(1-\beta). \quad (29)$$

The intrinsic dissociation rate constant is given by

$$k_d^{\mathrm{AviaB}} = \Phi\frac{P(\sigma \mid A) + P^{\mathrm{hop}}(B \mid A)P^{\mathrm{norm}}(\sigma \mid B)}{1-\alpha\beta P^{\mathrm{norm}}(\sigma \mid B)}. \quad (30)$$

As all ingredients follow directly from the FFS (or TIS) calculation, this completes the description of the dissociation rate constant.

2.4. The effective association rate constant

In this section, we will give the expression of the effective association. The on-rate at the σ surface is

$$k_{\mathrm{on}}^{\mathrm{AviaB}}(\sigma) = k_{\mathrm{D}}(\sigma)P_{\mathrm{A}}(\sigma), \tag{31}$$

where $k_{\mathrm{D}}(\sigma)$, the diffusion-limited reaction rate constant, is the rate at which the particles arrive at the σ-interface from infinity, i.e. the bulk.

The probability $P_{\mathrm{A}}(\sigma)$ is the probability to reach A when initialised at σ. This includes all paths with B as intermediate. This means that the paths via state B are (temporarily) trapped, but can hop to A, or can escape and rebind to B, or can escape to σ. As for dissociation discussed above, we do not want to compute that explicitly, as B is potentially a strongly bound state, and trajectory times can be very long. However, we can again decompose the paths into different categories, each with their own probability (see Figure 2): (1) direct paths to A, with probability $P(A \mid \sigma)$, (2) indirect paths to B with probability $P(B \mid \sigma)$, (3) rebinding paths from B to B with probability $P(B \mid B)$, (4) paths that leave B and reach sigma with probability $P(\sigma \mid B)$, (5) hopping paths from B to A with probability $P_{\mathrm{hop}}(A \mid B)$.

Now, these probabilities are not necessarily always the same as the ones for the dissociation. However, when the distribution of particles and orientations is isotropic at the σ surface, we can use the same expressions as for the dissociation. In the simulations, it is thus important to check that the distributions have become isotropic at the surface σ [15].

The probability $P_{\mathrm{A}}(\sigma)$ consists of direct paths and indirect paths via B. The indirect paths that visit B, then rebind an arbitrary number of times to B before going to A, have a probability

$$\begin{aligned} P^{\sigma\mathrm{BA}}(A \mid \sigma) &= P(B \mid \sigma)P_{\mathrm{hop}}(A \mid B) + P(B \mid \sigma)p(B \mid B) \\ &\quad \times P_{\mathrm{hop}}(A \mid B) + P(B \mid \sigma)P(B \mid B)^2 \\ &\quad \times P_{\mathrm{hop}}(A \mid B) + \cdots \end{aligned} \tag{32}$$

This can be re-summed to give

$$P^{\sigma\mathrm{BA}}(A \mid \sigma) = P(B \mid \sigma)P_{\mathrm{hop}}(A \mid B)\frac{1}{1 - P(B \mid B)} \tag{33}$$

$$\begin{aligned} &= P(B \mid \sigma)\frac{P_{\mathrm{hop}}(A \mid B)}{P(\sigma \mid B) + P_{\mathrm{hop}}(A \mid B)} \\ &\equiv P(B \mid \sigma)P_{\mathrm{hop}}^{\mathrm{norm}}(A \mid B), \end{aligned} \tag{34}$$

where in the last equality we defined the normalised hopping probability $P_{\mathrm{hop}}^{\mathrm{norm}}(A \mid B)$ to reach σ from B in an arbitrary number of rebinding steps.

Using these above probabilities, the probability $P_A(\sigma)$ is now

$$\begin{aligned} P_{\mathrm{A}}(\sigma) &= P(A \mid \sigma) + P(A \mid \sigma)P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma) \\ &\quad + P(A \mid \sigma)P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma)^2 \cdots \\ &\quad + P^{\sigma\mathrm{BA}}(A \mid \sigma) + P^{\sigma\mathrm{BA}}(A \mid \sigma)P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma) \\ &\quad + P^{\sigma\mathrm{BA}}(A \mid \sigma)P^{\sigma\mathrm{B}\sigma}(\sigma \mid \sigma)^2 \cdots \end{aligned} \tag{35}$$

This can be re-summed to give

$$\begin{aligned} P_{\mathrm{A}}(\sigma) &= \frac{P(A \mid \sigma) + P^{\sigma\mathrm{BA}}}{1 - P^{\sigma\mathrm{B}\sigma}} \\ &= \frac{P(A \mid \sigma) + P(B \mid \sigma)P_{\mathrm{hop}}^{\mathrm{norm}}(A \mid B)}{1 - P(B \mid \sigma)P^{\mathrm{norm}}(\sigma \mid B)}. \end{aligned} \tag{36}$$

The effective association rate is obtained by using Equation (36) in Equation (31),

$$k_{\mathrm{on}}^{\mathrm{AviaB}} = k_{\mathrm{D}}(\sigma)\frac{P(A \mid \sigma) + P(B \mid \sigma)P_{\mathrm{hop}}^{\mathrm{norm}}(A \mid B)}{1 - P(B \mid \sigma)P^{\mathrm{norm}}(\sigma \mid B)}. \tag{37}$$

Using Equation (28) and defining $\beta = 1 - P(\infty \mid \sigma)$, the final explicit expression for the effective on-rate becomes

$$k_{\mathrm{on}}^{\mathrm{AviaB}} = k_{\mathrm{D}}(\sigma)\frac{(1-\alpha)\beta + \alpha\beta P_{\mathrm{hop}}^{\mathrm{norm}}(A \mid B)}{1 - \alpha\beta P^{\mathrm{norm}}(\sigma \mid B)}. \tag{38}$$

As all ingredients follow directly from the FFS (or TIS) calculation, this completes the description of the rate constants. We note that with $k_{\mathrm{on}}^{\mathrm{AviaB}}$ given by Equation (38) and k_D given by $k_D(\sigma) = 4\pi\sigma D$, the intrinsic association rate constant k_a can directly be derived from

$$\frac{1}{k_{\mathrm{on}}^{\mathrm{AviaB}}} = \frac{1}{k_{\mathrm{a}}^{\mathrm{AviaB}}(\sigma)} + \frac{1}{k_{\mathrm{D}}(\sigma)}. \tag{39}$$

2.5. Comparison with previous work

In [28] we derived microscopic expressions for the rate constants for association to and dissociation from patch A, in the case of two target patches A and B:

$$k_{\mathrm{on}}^{A} = k_{\mathrm{D}}(\sigma)(1-\alpha)(1 - P(\infty \mid \sigma)), \tag{40}$$

$$k_{\mathrm{off}}^{A} = \Phi P(\sigma \mid r_0^A)P(\infty \mid \sigma). \tag{41}$$

Here α is the splitting probability to go to B rather than A, as in the previous sections.

Since we have now generalised Equations (38) and (29) for the effective association and dissociation rates, respectively, Equations (40) and (41) should follow from these. To see the connection, we assume that when B is

considered a target state, all subsequent transitions out of B should be ignored in the rate calculations. This is can be achieved by setting the probability $P(A \mid B) = 0$ and $P(\sigma \mid B) = 0$. Equations (38) and (29) then immediately reduce to Equations (40) and (41), respectively.

Finally, in [28] we showed that the association and dissociation rate constants for the scenario with only a *single* patch A is also given by Equations (40) and (41), but with $\alpha = 0$ for the association rate constant (because there is no B site). We indeed find that our principal results for the association and dissociation rate constants, Equations (38) and (29), respectively, reduce to the same expressions when α is set to zero in Equations (38) and (29)

$$k_{\text{on}}^{A} = k_{\text{D}}(\sigma)(1 - P(\infty \mid \sigma)), \tag{42}$$

$$k_{\text{off}}^{A} = \Phi_0 P(\sigma \mid r_0^A) P(\infty \mid \sigma), \tag{43}$$

which indeed are identical to Equations (40) and (41), for $\alpha = 0$.

2.6. The link with transition path theory

Instead of the treatment detailed above, using conditional probabilities as obtained from FFS or TIS, one can also obtain the association and dissociation rates from the rates for the individual transitions between A, B, σ and U. The overall rates of association and dissociation between A and U can then be calculated via transition path theory (TPT) [29–33]. In TPT analysis, the commitment probability, or committor, is calculated from the transition matrix, $\mathbf{T} = \exp(\mathbf{K}\tau)$, where T_{ij} is the transition probability between i and j within a certain lag time τ. This transition matrix can be computed from the rate matrix K, with k_{ij} the rate constant between i and j.

If we consider the dissociation process, the committor q_i^+ is the probability the final state U is reached from state i before returning to state A and can be obtained in general by solving the linear set of equations [29]:

$$-q_i^+ + \sum_{j \in I} T_{ij} q_j^+ = -\sum_{j \in U} T_{ij}, \tag{44}$$

with $q_A^+ = 0$ and $q_U^+ = 1$. States I denote all intermediate states.

For the problem discussed in this work, the overall dissociation rate constant is given by [29,33]

$$k_{AU} = \frac{\sum_{J \neq A} \pi_A T_{AJ} q_J^+}{\tau \sum_J \pi_J q_J^-}, \tag{45}$$

with $q^- = 1 - q^+$, and π_J is the equilibrium population of state J. When the process of leaving A is a rare event, the denominator is effectively π_A, as all intermediate states are supposed to be barely populated, and $q_U^- = 0$. Hence Equation (45) reduces for our system to

$$k_{AU} = \frac{1}{\tau \pi_A} \sum_{J \neq A} \pi_A T_{AJ} q_J^+ = \frac{1}{\tau} T_{AB} q_B^+ + T_{A\sigma} q_\sigma^+. \tag{46}$$

To make the connection to our current work we exploit the fact that for the dissociation process $A \to U$, the following identification can be made:

$$T_{AJ}/\tau = \Phi P(J \mid A). \tag{47}$$

Combining this identification with Equation (46) then yields

$$k_{AU} = \Phi[P(B \mid A) q_B^+ + P(\sigma \mid A) q_\sigma^+]. \tag{48}$$

The committor q_B^+ is

$$q_B^+ = \frac{P(\sigma \mid B)}{P(\sigma \mid B) + P(A \mid B)} q_\sigma^+ = P^{\text{norm}}(\sigma \mid B) q_\sigma^+, \tag{49}$$

while the committor q_σ^+ is

$$q_\sigma^+ = \frac{P(\infty \mid \sigma)}{1 - P^{\sigma B \sigma}(\sigma \mid \sigma)}, \tag{50}$$

resulting in

$$k_{AU} = \Phi \left[P(B \mid A) P^{\text{norm}}(\sigma \mid B) + P(\sigma \mid A) \right] \times \frac{P(\infty \mid \sigma)}{1 - P^{\sigma B \sigma}(\sigma \mid \sigma)}, \tag{51}$$

which is indeed identical to Equation (11), the effective dissociation rate before correction. As similar derivation can be done for the association rate constant resulting in expression identical to Equation (37).

Note that in [32–34] we used the TPT expressions to compute effective association and dissociation rates. In this work, we achieve the same by carefully accounting for the different trajectories, as explained in the previous sections. The large advantage of this treatment is that we only have to conduct one FFS or TIS for (each of) the dissociation processes.

Moreover, the TPT approach will not yield the desired effective rates, as it does not directly allow the correction for the use of the r_n surface (instead it yields the rates corresponding to the finite simulation system). However, as such correction only acts on the $P(\infty \mid \sigma)$ term, one could in principle apply this correction also to the corresponding terms in TPT. This would also allow the treatment of multiple intermediate states, something that can become cumbersome in an explicit derivation.

3. Results and discussion

Our model system consists of two particles that are spherical with an isotropic centre-of-mass interaction, dressed with one (for the enzyme) or two (for the substrate) sticky spots on their surface called 'patches', which allow for highly directional, anisotropic interactions. The enzyme–substrate pair experiences a strong attractive potential, $U_{\rm s}(r)$, over a narrow band of orientations. This specific attraction depends on the distance, r, between the patches, i.e. stronger attraction when the patches are closer. When the patchy particles approach each other, they experience an isotropic repulsive potential, $U_{\rm rep}(R)$, which is a function of the centre-of-mass distance, R. In addition, particles experience a weak, isotropic, non-specific attraction, $U_{\rm ns}(R)$. The total patch potential reads:

$$U_{\rm an}(R,r) = \sum_{\rm i=1}^{n} U_{\rm s}(r_{\rm i}) + U_{\rm rep}(R) + U_{\rm ns}(R), \qquad (52)$$

where n is the number of patches on the substrate (two in the context of this paper) and $r_{\rm i}$ are the inter-patch distances between the patch of the enzyme and the ith patch of the substrate. $U_{\rm s}(r)$, $U_{\rm rep}(R)$ and $U_{\rm ns}(R)$ have the form

$$U_i(x) = \begin{cases} \epsilon_i \left(1 - a_i \left(\dfrac{x}{\sigma_{\rm an}}\right)^2\right) & \text{if } x < x_i^\star, \\ \epsilon_i b_i \left(\dfrac{x_i^c}{\sigma_{\rm an}} - \dfrac{x}{\sigma_{\rm an}}\right)^2 & \text{if } x_i^\star < x < x_i^c, \\ 0 & \text{otherwise,} \end{cases} \qquad (53)$$

with $i = \{{\rm s}, {\rm rep}, {\rm ns}\}$ respectively. The overall strength ϵ_i, the length scale $\sigma_{\rm an} = 5\,{\rm nm}$, the stiffness a_i and the parameter $x_i^\star$, which combined with a_i determines the range of the potential, are free parameters. Cut-offs x_i^c and smoothing parameters b_i are fixed by requiring continuity and differentiability at $x_i^\star$. In this paper, we use the same parameter settings as in [14,15,35]: $\epsilon_{\rm s} = 20 k_{\rm B}T$, $a_{\rm s} = 20$ and $r_{\rm att}^* = 0.1\sigma_{\rm an}$, implying $b_{\rm s} = 5$ and $r_{\rm s}^c = 0.5\sigma_{\rm an}$; $\epsilon_{\rm rep} = 100 k_{\rm BT}$, $a_{\rm rep} = 1$ and $R_{\rm rep}^* = 0.85\sigma_{\rm an}$, implying $b_{\rm rep} = 2.6036$; and $R_{\rm rep}^c = 1.1764\sigma_{\rm an}$; and $a_{\rm ns} = 1$ and $R_{\rm ns}^* = 0.85\sigma_{\rm an}$, implying $b_{\rm ns} = 2.6036$; and $R_{\rm ns}^c = 1.1764\sigma_{\rm an}$. $\epsilon_{\rm ns}$ is varied from $2k_{\rm B}T$ to $20k_{\rm B}T$ with steps of $2k_{\rm B}T$. When the patches are aligned ($r = R - \sigma_{\rm an}$) and misaligned ($r = R + \sigma_{\rm an}$). When the patches are aligned, particles experience both specific and non-specific attraction, creating a deeper potential well and a stronger bond. When the patches are misaligned, $U_{\rm s} = 0$ and the particles only experience the weak $U_{\rm ns}$ which results in a shallow potential well and a weaker bond. The non-specific attraction, however, promotes realignment since the particles do not diffuse away immediately.

These parameters were chosen to mimic protein association with dissociation constants on the order of nM to mM [36–38]. The length scale $\sigma_{an} = 5\,{\rm nm}$ represents a typical diameter for a protein substrate or enzyme. The absolute rate constants, being on the millisecond–second timescale, as well as the equilibrium constant $K_{\rm eq}$, given by the ratio of the dissociation and association rate constants, are typical for biological systems [36–38].

All simulations are performed using a Brownian dynamics integrator [39]. The system-specific parameters of the simulation are as follows: the time step $\delta t = 0.1\,{\rm ns}$, the mass of the particle is $m = 50\,{\rm kDa}$, the mass moment of inertia $M = (8/15)m\sigma^2$, the translational and rotational friction coefficients are $\gamma = k_{\rm B}T/D_t m$ and $\Gamma = k_{\rm B}T/D_r M$, respectively, where $D_t = 1\,{\rm um}^2/{\rm s}$ and D_r translational and rotational diffusion constants, respectively, $k_{\rm B} = 1.38 \times 10^{-23}\,{\rm JK}^{-1}$ is the Boltzmann constant and $T = 300\,{\rm K}$ is the temperature of the system.

As we employ an anisotropic interaction potential, a geometrical definition of interfaces for FFS is not straightforward. Therefore, we base the interfaces on the potential energy up to the potential cut-off. The first interface (r_0-interface), which defines the bound state, is located at $18k_BT$. The other interfaces are located at 15, 10, and 5 k_BT. The interface at the cut-off of the potential is defined by zero energy and $R = 1.6\sigma_{an}$. Beyond the cut-off of the potential, the interfaces are defined by the interparticle distance $R = 1.7, 1.9, 2.1, 2.3, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5$. Finally the r_n-interface is located at $R = 7\sigma_{an}$.

We conducted multiple FFS simulations as described in the method sections for several values of the rotational diffusion constant D_r. For simplicity we kept the same system as described in [28], which had identical patches A and B. We performed simulations for two systems, one where the patches are separated by an angle $\theta = 90^\circ$ and one where this angle is $\theta = 120^\circ$. Using the generalised FFS expression introduced in [28], we extracted from these simulations the conditional probabilities $P(\sigma \mid A), P(A \mid B), \alpha, \beta = 1 - P(\infty \mid \sigma) = (1 - P(r_{\rm n} \mid \sigma))/(1 - P(r_{\rm n} \mid \sigma)\Omega)$ and the flux through the first interface Φ. These values are reported in Table 1. Since $P(\sigma \mid B) = P(\sigma \mid A)$ and $P(B \mid A) = P(A \mid B)$, we do not report these numbers explicitly.

From these numbers the probabilities $P(B \mid \sigma) = \alpha\beta$ and $P(A \mid \sigma) = (1 - \alpha)\beta$ were computed, which in turn led to values for $P^{\rm norm}(\sigma \mid B)$ and $P_{\rm hop}^{\rm norm}(A \mid B)$ (see Equations (6) and (34), and Table 1 for the numbers), which finally yielded the intrinsic and effective association and dissociation constants, $k_{\rm on}^{\rm AviaB}$, $k_a^{\rm AviaB}$, $k_{\rm off}^{\rm AviaB}$ and $k_d^{\rm AviaB}$ from Equations (38), (39), (29) and (30), respectively. We also computed the relevant rate

Table 1. All components going into the rate constant expressions for several values of D_r for $\theta = 90^\circ$ and $\theta = 120^\circ$.

D_r (rad^2/s)	$P(\sigma \mid A)$	$P(A \mid B)$	$P^{\text{norm}}(\sigma \mid B)$	$P^{\text{norm}}_{\text{hop}}(A \mid B)$	β	ϕ (s^{-1})	α	$1 - P^{\sigma B \sigma}$
$\theta = 90^\circ$								
56,000	5.27×10^{-8}	4.22×10^{-8}	0.554	0.446	0.352	1.19×10^{7}	0.5	0.902
64,000	5.10×10^{-8}	4.61×10^{-8}	0.525	0.475	0.358	1.23×10^{7}	0.5	0.906
72,000	4.89×10^{-8}	4.48×10^{-8}	0.519	0.481	0.358	1.26×10^{7}	0.5	0.907
80,000	5.20×10^{-8}	4.93×10^{-8}	0.514	0.486	0.368	1.30×10^{7}	0.5	0.906
88,000	5.92×10^{-8}	5.90×10^{-8}	0.501	0.499	0.362	1.33×10^{7}	0.5	0.909
$\theta = 120^\circ$								
56,000	5.28×10^{-8}	3.58×10^{-8}	0.595	0.405	0.364	1.19×10^{7}	0.5	0.892
64,000	5.24×10^{-8}	3.84×10^{-8}	0.577	0.423	0.365	1.23×10^{7}	0.5	0.895
72,000	5.25×10^{-8}	3.96×10^{-8}	0.570	0.430	0.368	1.26×10^{7}	0.5	0.895
80,000	5.45×10^{-8}	4.33×10^{-8}	0.557	0.443	0.370	1.30×10^{7}	0.5	0.897
88,000	5.93×10^{-8}	5.00×10^{-8}	0.543	0.457	0.371	1.33×10^{7}	0.5	0.899

Note: The label $1 - P^{\sigma B \sigma}$ refers to the denominators in Equations (38) and (29), which account for the paths from σ to σ via B.

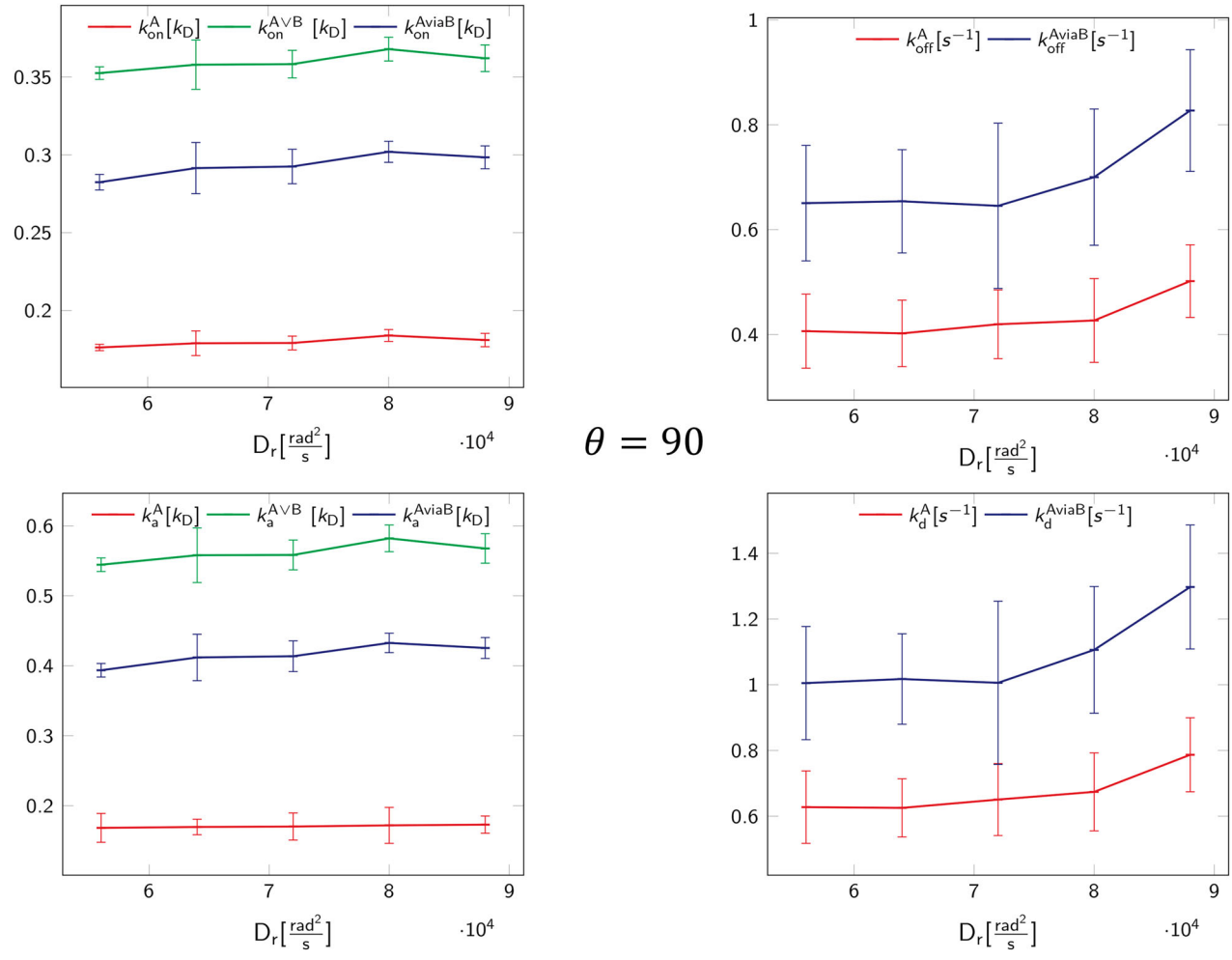


Figure 3. Intrinsic and effective association and dissociation rates for $\theta = 90^\circ$.

constants for the (un)binding to a substrate with a single patch A, k_a^A, k_d^A, k_{on}^A, k_{off}^A, and for association to either of the two patches of a substrate with two patches, $k_a^{A\vee B}$ and $k_{\text{on}}^{A\vee B}$. All these rate constants are shown in Figures 3 and 4 for both $\theta = 90^\circ$ and $\theta = 120^\circ$, respectively, and reported in Table 2.

The first thing to note is that the effective association rate $k_{\text{on}}^{\text{AviaB}}$ is always in between the values of k_{on}^A and $k_{\text{on}}^{A\vee B}$. That is, $k_{\text{on}}^{\text{AviaB}}$ is higher than the single patch value, which is to be expected, but lower than the 'either A or B' rate constant, due to pathways that leave B again and then get lost into the bulk before reaching A. The contribution of the paths that reach A via B is about 60% of the paths that reach A directly. We observe similar behaviour for k_a. The dissociation rates are affected in the same way, as expected from detailed balance.

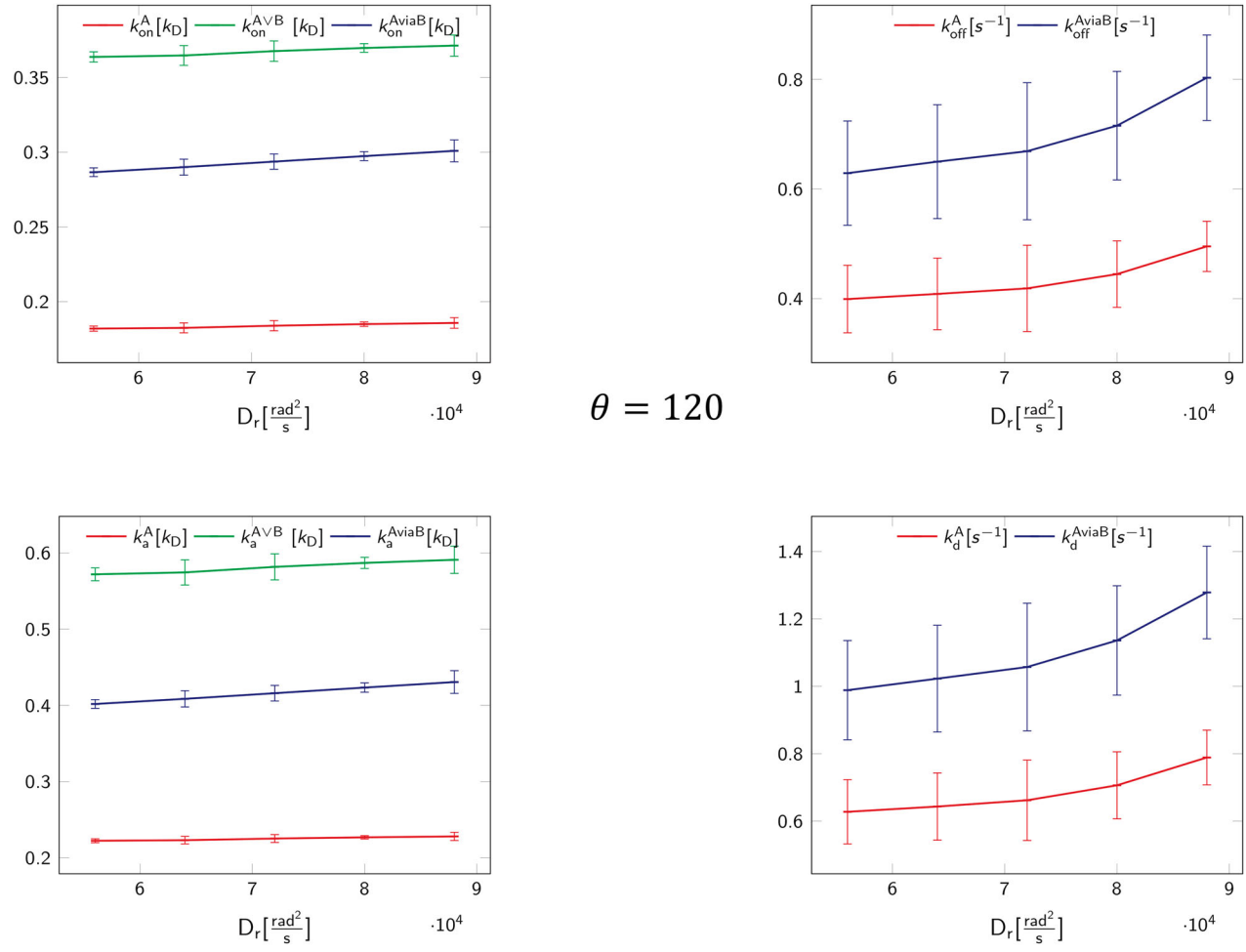


Figure 4. Intrinsic and effective association and dissociation rates for $\theta = 120^\circ$.

Table 2. All computed rate constants for several values of D_r for $\theta = 90^\circ$ and $\theta = 120^\circ$.

D_r (rad^2/s)	k_{on}^{AviaB}	$k_{on}^{A\vee B}$	k_{on}^{A}	k_{off}^{AviaB}	k_{off}^{A}	k_{a}^{AviaB}	$k_{a}^{A\vee B}$	k_{a}^{A}	k_d	k_{d}^{A}	K_{eq}^{AviaB}	K_{eq}^{A}
$\theta = 90^\circ$												
56,000	0.282	0.352	0.176	0.651	0.406	0.394	0.544	0.214	1.005	0.628	2.305	2.305
64,000	0.291	0.358	0.179	0.654	0.402	0.412	0.558	0.218	1.017	0.625	2.257	2.257
72,000	0.292	0.358	0.179	0.645	0.396	0.414	0.558	0.218	1.006	0.618	2.213	2.213
80,000	0.302	0.368	0.184	0.700	0.427	0.433	0.582	0.225	1.106	0.674	2.326	2.326
88,000	0.298	0.362	0.181	0.827	0.502	0.425	0.568	0.221	1.297	0.787	2.771	2.771
$\theta = 120^\circ$												
56,000	0.287	0.364	0.182	0.629	0.399	0.402	0.572	0.222	0.988	0.628	2.196	2.196
64,000	0.290	0.365	0.182	0.650	0.409	0.409	0.574	0.223	1.023	0.643	2.243	2.243
72,000	0.294	0.368	0.184	0.669	0.419	0.416	0.582	0.225	1.057	0.662	2.282	2.282
80,000	0.297	0.370	0.185	0.716	0.445	0.423	0.587	0.227	1.136	0.706	2.404	2.404
88,000	0.301	0.371	0.186	0.803	0.495	0.431	0.591	0.228	1.278	0.789	2.665	2.665

Notes: All association rate constants are in units of k_D, while all dissociation rate constants are in units of s^{-1}. The last two columns report the equilibrium constants in units of $(sk_D)^{-1}$, computed by the ratio of the on and off rates.

Finally there is very little difference between $\theta = 90^\circ$ and $\theta = 120^\circ$. At first sight this seems to be caused by the fact that the patches are too far away to influence each other. However, when zooming in on the hopping probability $P(A \mid B)$ and normalised splitting probabilities $P_{\text{hop}}^{\text{norm}}(A \mid B)$ and $P^{\text{norm}}(\sigma \mid B)$, there is a substantial difference between $\theta = 90^\circ$ and $\theta = 120^\circ$ (see Table 1). As an example, for $D_r = 88000$; the hopping probability $P(A \mid B)$ increases by almost 20% when lowering the angle from $\theta = 120^\circ$ to $\theta = 90^\circ$. Yet, the effective rates are almost independent of the angle. This indicates that the increase in the rate due to higher $P_{\text{hop}}^{\text{norm}}(A \mid B)$ for the $\theta = 90^\circ$ is mostly compensated by the slightly higher rebinding correction $1 - P^{\sigma B \sigma}$, and the slight reduction

in β. We expect for smaller θ the effective rates to go up, until eventually it reaches the maximum value of the 'either A or B' case.

In the last two columns of Table 2, we give the equilibrium constant computed for the AviaB and the single patch cases. While they vary a bit with D_r due to statistical uncertainty, entries in both columns are identical, as it should be, thus providing a sanity check.

In our framework, we implicitly assume that association and dissociation obeys Markovian two-state kinetics, characterised by exponential relaxation with well-defined rate constants. However, the presence of the intermediate patch B could lead to marked deviations from exponential relaxation. Our simulations were done for identical patches, which violates the assumption $K_{\mathrm{eq}}^{A} \ll K_{\mathrm{eq}}^{B}$. So, in fact, the system simulated here probably cannot be described by effective two-state kinetics, but instead will exhibit three-state kinetics. Therefore, our numerical results in terms of effective rate constants should be seen as a proof of principle, rather than a direct experimentally relevant result. While we do not explicitly study the dependence on the relative strength of the patches here, we argue that increasing the strength of patch A significantly, e.g. by adding a deep and narrow potential located at the centre of patch A, would increase the dwell time in A significantly. This in turn would make patch B metastable w.r.t. to patch A, and the system would revert to two-state kinetics, with well-defined effective association and dissociation rate constants. While this will change the results quantitatively, because the hopping and escape probabilities will change, the qualitative results, namely the fact that $k_{\mathrm{on}}^{\mathrm{AviaB}}$ is in between $k_{\mathrm{on}}^{\mathrm{A}}$ and $k_{\mathrm{on}}^{\mathrm{A}\vee\mathrm{B}}$, and that the association and dissociation rates are substantially influenced by the intermediate patch, will likely also be valid for unequal patches for which $K_{\mathrm{eq}}^{A} \ll K_{\mathrm{eq}}^{B}$.

More generally, the question whether the system can be described as a two-state system depends on the details of the system and needs to be assessed with care. In the extreme scenario where the rate of hopping from A to B is much lower than that of escaping from A to the unbound state, dissociation would be effectively described by two-state state kinetics. Also in the other extreme, where the hopping rate from A to B is much higher compared to escaping from A to the unbound state, two-state kinetics would prevail. However, when the rate of hopping becomes comparable to the escape rate, then a two-state description probably breaks down (see also the Supplementary Material of [40] elsewhere in this Special Edition). Indeed, it would be of interest to study systematically under which conditions the binding to a patch in the presence of other patches can still be described via two-state kinetics. We leave this for future work.

4. Conclusions

In this paper, we derived microscopic expressions for the intrinsic and effective rate constants for association and dissociation of enzymes to substrates with two patches, one of which can be considered an intermediate state. In that case the (un)binding process follows two-state kinetics. The derived expressions can be used in conjunction with rare-event simulations of the dissociation processes. The final expressions for the effective on and off rates contain contributions of direct and indirect trajectories via the intermediate patch, and account for rebinding events. We illustrated the use of the new expressions by applying it to the same enzyme–substrate system discussed in [28]. We computed the rate constants as a function of the rotational diffusion constant, D_{r}, and the spacing between the two patches on the surface of the substrate. We find that the association and dissociate rate is dramatically affected by the inclusion of the intermediate state. Leaving this state out, even for a strongly binding patch, yields an error of more than 50%. For large patches this is expected to be even more so.

Evaluation of these rate constants is useful for understanding in general the association reactions in an enzyme–substrate system and to study the response characteristics of such a system. The intrinsic rate constants also serve as input parameters for a multi-scale simulation [35,41], where, by using these rate constants, explicit simulations of the association reactions can be avoided, which dramatically speeds up the simulations.

More generally, the expressions derived in this paper can be used for simulation of all processes in which association and dissociation plays a role, not only chemical reactions, but also self-assembly of, for instance, colloidal particles.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is part of the Industrial Partnership Programme (IPP) 'Computational sciences for energy research' of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organization for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V.

References

[1] R. Hermsen, S. Tans and P.R. ten Wolde, PLoS Comput. Biol. **2** (12), e164 (2006).
[2] C.C. Govern and P.R. ten Wolde, Proc. Nat. Acad. Sci. USA **111** (49), 17486 (2014).

[3] J.S. van Zon, M.J. Morelli and P.R. ten Wolde, Biophys. J. **91** (12), 4350 (2006).
[4] O.G. Berg, R.B. Winter and P.H. Von Hippel, Biochemistry **20** (24), 6929 (1981).
[5] V. Vanguri, C.C. Govern, R. Smith and E.S. Huseby, Proc. Nat. Acad. Sci. USA **110** (1), 288 (2013).
[6] N. Agmon and A. Szabo, J. Chem. Phys. **92**, 5270 (1990).
[7] O.G. Berg and P.H. von Hippel, Annu. Rev. Biophys. Chem. **14** (1), 131 (1985).
[8] S.H. Northrup and H.P. Erickson, Proc. Nat. Acad. Sci. USA **89** (8), 3338 (1992).
[9] H.X. Zhou, J. Chem. Phys. **108** (19), 8139 (1998).
[10] T.S. van Erp, D. Moroni and P.G. Bolhuis, J. Chem. Phys. **118** (17), 7762 (2003).
[11] R. Cabriolu, K.M.S. Refsnes, P.G. Bolhuis and T.S. van Erp, J. Chem. Phys. **147** (15), 152722 (2017).
[12] R.J. Allen, P.B. Warren and P.R. ten Wolde, Phys. Rev. Lett. **94** (1), 018104 (2005).
[13] R.J. Allen, C. Valeriani, P.R. ten Wolde, J. Phys. Cond. Matt. **21** (46), 463102 (2009).
[14] A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde, Faraday Discuss. **195**, 421 (2016).
[15] A. Vijaykumar, P.R. ten Wolde and P.G. Bolhuis, J. Chem. Phys. **147** (18), 184108 (2017).
[16] K. Takahashi, S. Tănase-Nicola and P.R. ten Wolde, Proc. Nat. Acad. Sci. USA **107** (6), 2473 (2010).
[17] J. Paijmans, D.K. Lubensky and P.R. ten Wolde, PLoS Comput. Biol. **13** (3), e1005415 (2017).
[18] L.F. Larrondo, C. Olivares-Yanez, C.L. Baker, J.J. Loros and J.C. Dunlap, Science **347** (6221), 1257277 (2015).
[19] I. Edery, L.J. Zwiebel, M.E. Dembinska and M. Rosbash, Proc. Nat. Acad. Sci. USA **91** (6), 2260 (1994).
[20] G.R. Crabtree and E.N. Olson, Cell **109** (Suppl), S67 (2002).
[21] P. Miller, A.M. Zhabotinsky, J.E. Lisman and X.J. Wang, PLoS Biol. **3** (4), e107 (2005).
[22] C.Y. Huang and J.E. Ferrell Jr, Proc. Nat. Acad. Sci. USA **93**, 10078 (1996).
[23] J.E. Ferrell Jr, Trends Biochem. Sci. **21**, 460 (1996).
[24] J. Gunawardena, Proc. Natl. Acad. Sci. USA **102** (41), 14617 (2005).
[25] N.I. Markevich, J.B. Hoek and B.N. Kholodenko, J. Cell Biol. **164** (3), 353 (2004).
[26] K. Aokia, M. Yamada, K. Kunida, S. Yasuda and M. Matsuda, Proc. Nat. Acad. Sci. USA **108** (31), 12675 (2011).
[27] I.V. Gopich and A. Szabo, Proc. Nat. Acad. Sci. USA **110** (49), 19784 (2013).
[28] A. Vijaykumar, P.R. ten Wolde, P.G. Bolhuis, J. Chem. Phys. **148** (12), 124109 (2018).
[29] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T.R. Weikl, Proc. Nat. Acad. Sci. USA **106** (45), 19011 (2009).
[30] E. Weinan and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61** (1), 391 (2010).
[31] M.R. Perkett and M.F. Hagan, J. Chem. Phys. **140** (21), 214101 (2014).
[32] A.C. Newton, J. Groenewold, W.K. Kegel and P.G. Bolhuis, J. Chem. Phys. **146** (23), 234901 (2017).
[33] A.C. Newton, R. Kools, D.W.H. Swenson and P.G. Bolhuis, J. Chem. Phys. **147** (15), 155101 (2017).
[34] A.C. Newton, J. Groenewold, W.K. Kegel and P.G. Bolhuis, Proc. Nat. Acad. Sci. **112** (50), 15308 (2015).
[35] A. Vijaykumar, T.E. Ouldridge, P.R. ten Wolde and P.G. Bolhuis, J. Chem. Phys. **146** (11), 114106 (2017).
[36] G. Schreiber, G. Haran and H.X. Zhou, Chem. Rev. **109** (3), 839 (2009).
[37] R. Milo, P. Jorgensen, U. Moran, G. Weber and M. Springer, Nucl. Acids Res. **38** (suppl_1), D750 (2009).
[38] A database of useful biological numbers can be found at URL http://bionumbers.hms.harvard.edu 2018).
[39] R.L. Davidchack, T.E. Ouldridge and M.V. Tretyakov, J. Chem. Phys. **142** (14), 144114 (2015).
[40] T.C.T. Michaels, L.X. Liu, S. Curk, P.G. Bolhuis, A. Saric, and T.P.J. Knowles, Mol. Phys. (In press).
[41] A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde, J. Chem. Phys. **143** (21), 214102 (2015).