Multi-Scale simulation of reaction-diffusion systems

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 24 oktober 2017, te 14:00 uur

door

Adithya Vijaykumar

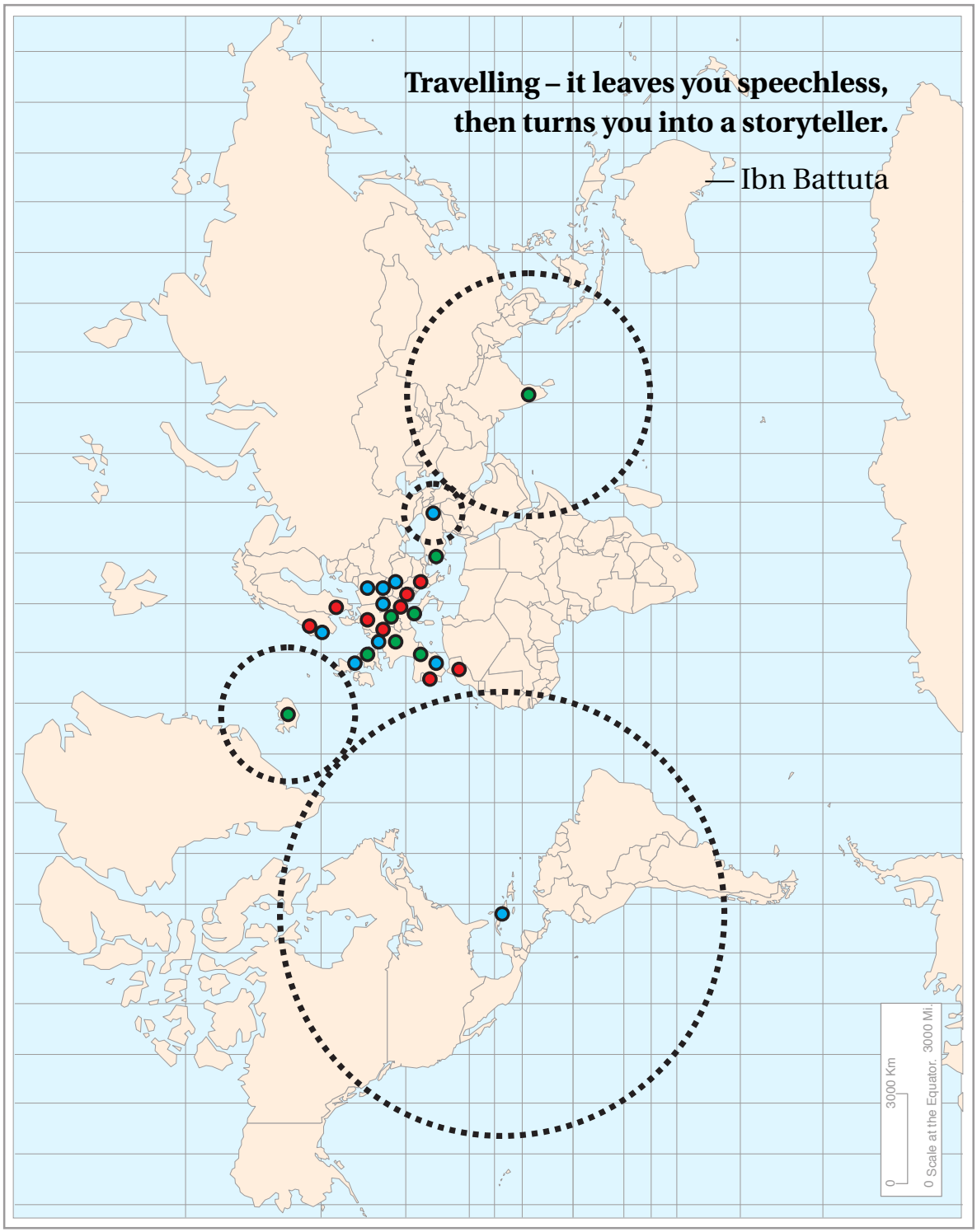
geboren te Bangalore, India.

Promotiecommissie:

Promotors:	Prof. dr. P.G. Bolhuis *Universiteit van Amsterdam* Prof. dr. P. R. Ten Wolde *Vrije Universiteit*
Overige Leden:	Prof. dr. J. Elf *Uppsala University* Prof. dr. T. Shimizu *Vrije Universiteit* Prof. dr. S. Woutersen *Universiteit van Amsterdam* Prof. dr. E. J. Meijer *Universiteit van Amsterdam* Dr. T. E. Ouldridge *Imperial College London* Dr. ir. B. Ensing *Universiteit van Amsterdam*

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

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.


Travelling – it leaves you speechless,
then turns you into a storyteller.
— Ibn Battuta
0
3000 Km
0 Scale at the Equator. 3000 Mi.

Cover Illustration: The painting on the front is inspired by Mondriaan's style of neoplasticism. It tries to capture a ship sailing on the sea during sunset. The sea and the sun and the sky does not require a lot of detail, hence the use of larger sized grids in the outer parts. To capture the fine details of the ship, a smaller grid is necessary. This constitutes a composition made of multiple grid sizes. This is analogous to the simulation technique described in the thesis. When the particles are far apart, we use a mesoscopic technique GFRD, that captures the larger length and time scales associated with diffusion, and when the particles come closer, we switch to a more detailed microscopic technique to capture the dynamics of interaction between the particles. The back cover is an abstract depiction of patchy particles, that are used in the simulations.

Cover Design: Adithya Vijaykumar, 2017

Printed by: Ipskamp, The Netherlands

ISBN 9789402807974

CONTENTS

1 Introduction **1**
1.1 Algorithms to simulate biochemical networks . . . 2
1.2 Overview of the thesis . . . 4
Bibliography . . . 6

2 Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations. **9**
2.1 Introduction . . . 10
2.2 The multi-scale scheme . . . 12
2.3 Results and discussion . . . 24
2.4 Conclusions and future work . . . 32
Bibliography . . . 33

3 Multiscale simulations of anisotropic particles combining Molecular Dynamics and Green's Function Reaction Dynamics **35**
3.1 Introduction . . . 36
3.2 Methods . . . 38
3.3 Results and discussion . . . 50
3.4 Conclusion . . . 57
3.5 Acknowledgements . . . 58
Bibliography . . . 58

4 The intrinsic rate constants in diffusion-influenced reactions **61**
4.1 Introduction . . . 62
4.2 Theory of diffusion-influenced reactions . . . 63
4.3 Effective positive flux expression . . . 64
4.4 Intrinsic dissociation rate and effective positive flux . . . 65
4.5 Computing the intrinsic rates for an interface at finite r_n . . . 67
4.6 Relation to other techniques . . . 70
4.7 Anisotropic interactions . . . 71
4.8 Conclusion . . . 76
Bibliography . . . 81

5 The magnitude of the intrinsic rate constant: How deep can association reactions be in the diffusion limited regime? **85**
5.1 Introduction . . . 86
5.2 Methods . . . 88
5.3 Results . . . 97
5.4 Discussion and Conclusions . . . 101
5.5 Acknowledgements . . . 103

Bibliography . . . 104

6 Transition rates for proteins associating to substrates with multiple binding sites using a novel generic Forward Flux Sampling expression 107
6.1 Introduction . . . 108
6.2 Methods . . . 109
6.3 Results . . . 117
6.4 Conclusion . . . 121
6.5 Acknowledgements . . . 122
Bibliography . . . 122

7 Multi-scale simulations of multi-site phosphorylation in a MAPK signalling pathway 125
7.1 Introduction . . . 126
7.2 System . . . 127
7.3 Methods . . . 130
7.4 System parameters . . . 134
7.5 Results . . . 134
7.6 Outlook . . . 135
Bibliography . . . 136

8 Summary 139

9 Samenvatting 141

10 List of Publications 143

11 Acknowledgements 145

1

INTRODUCTION

Have you ever wondered how a tiger gets its stripes? How a disease like ebola spreads? How a bacteria finds its centre before dividing? The behaviour of these systems is determined by the change in concentration of one or more of its components in space and time. These systems are called reaction-diffusion systems, where diffusing components react when they come in contact with each other. Reaction-diffusion systems are omnipresent in nature and in industry. In fact, a living cell can be thought of as a reaction-diffusion system.

Many reaction-diffusion processes span several orders in length and time scales. Examples are biochemical networks in living cells, catalytic reactions in, e.g. a fuel cell, or dynamics of surfactant/water/oil mixtures. In these systems, the spatial dynamics of the reactants at mesoscopic length and time scales of microns and seconds are coupled to the reactions between the molecules, at microscopic length and time scales of nanometers and milliseconds. This wide range of length and time scales makes these systems notoriously difficult to simulate.

The conventional strategy to simulate such reaction-diffusion systems, is to coarse-grain the dynamics at molecular scales. One approach is to divide the reaction volume into a number of subvolumes; particles can then react within one volume, but also diffuse from one volume to the next [1]. Another approach is to take into account the particulate nature of the reactants [2, 3, 4, 5, 6, 7]. In these particle-based schemes, coarse graining typically involves describing the reactants and their interactions by effective cross sections, diffusion constants and rate constants. This approach drastically coarse grains the dynamics at molecular scales. However, it is now clear that the dynamics at the molecular scale can qualitatively change the macroscopic behavior of the system. It has been shown that the rapid rebinding between a catalyst and its substrate can drastically change the rate of product formation[8], especially when one or both reactants are on a surface[9]. In such cases, the binding between substrate and enzyme, the conformational dynamics of the molecules, the formation and release of the product, the orientational diffusion of the reactants and products has to be described explicitly, and cannot be integrated out. Another approach, to simulate such systems is using (ab-initio)

molecular dynamics at all scales. This, while possible in principle, would be highly inefficient if not impossible in practice, since most CPU time would be spent on propagating the reactants towards each another before they can react. There is not a single technique, that can describe the dynamics at both the mesoscopic and the microscopic scales which is both accurate and efficient. These problems thus call for a particle-based, multi-scale modeling approach.

In this thesis we propose such a particle-based, multi-scale modelling approach. When far apart, the particles are propagated towards each other using a mesoscopic technique. Here, the dynamics at the molecular scales are not important since the particles are far apart. When the particles come close to each other, the algorithm switches to a microscopic/atomistic description of the particle dynamics, which takes into account the anisotropy of the particles and their conformational dynamics. In the multi-scale scheme proposed here, the particle dynamics at the different scales are thus described by different techniques, and the algorithm switches seamlessly between them based on predefined scenarios. The algorithm can be applied to any reaction-diffusion system, but it becomes particularly efficient when there is a clear separation in length and time scales between the dynamics at the microscopic and the mesoscopic levels. In the context of this thesis we use GFRD as the mesoscopic technique and BD as the microscopic technique, but other techniques like underdamped Langevin dynamics, or even Molecular dynamics, could in principle be used.

1.1. ALGORITHMS TO SIMULATE BIOCHEMICAL NETWORKS

A biochemical network is a network of chemical interactions and physical interactions between biomolecules, such as proteins and DNA in a cell. Biochemical networks are the computational devices of living cells. They allow cells to perform a number of computational tasks, such as signal integration and amplification.

The conventional approach to analyse biochemical networks is to numerically solve the macroscopic rate equations. The evolution of a network in this method is deterministic since the equations are ordinary differential equations. In this approach, the stochasticity can be neglected because it is assumed that the concentrations are high. The concentrations of molecules in a living cell however, can be very low, i.e. in the micromolar to nanomolar range. Because of this the stochasticity in biochemical networks cannot be neglected [10] [11].

The stochastic nature of biochemical networks can be addressed by adding a noise term to the macroscopic rate equations[12]. However, at low concentrations, the continuum approach is bound to fail as shown by Togashi and Kaneko [13] and Shnerb et al [14]. In a living cell, where the concentrations are low, we have to recognise the discrete nature of the reactants.

If the system is well-stirred and well-mixed at all times, the dynamics is *exactly* given by the chemical master equation.This description holds when there are many non-reactive collisions in between the reactive collisions, so that the system becomes self-stirred. This would mean that at each instant the particles are uniformly distributed in space. The accuracy of this description increases when the diffusion constant increases. If the system is linear, the CME can be solved analytically. This means that the variances, covariances

and power spectra can be obtained analytically and exactly. However, the reactions are often bimolecular, or even of higher order. In these cases the CME becomes non-linear and obtaining an analytical solution is not possible. However, there are approximate analytical solutions, such as the Linear-Noise Approximation. If, however, exact results are required for a non-linear CME, one typically has to resort to simulations. One technique is the SSA developed by Gillespie [15]. The SSA is an exact, event-driven kinetic Monte Carlo algorithm that generates trajectories which are consistent with the CME. The iterative nature of the Gillespie algorithm reduces memory demands, as opposed to when solving the CME directly. However, this algorithm only applies to the CME, which describes the dynamics of a well-stirred system, where space is integrated out. This is a serious limitation. A major reason for noise in biochemical networks is due to the fluctuations of various components is space [16]. It has also been shown that these fluctuations, which the mean-field analyses usually ignore, have a dramatic effect on the response of a signalling pathway[17]. Signals are often transmitted in biological systems by diffusion of particles. The concentration of these particles may not be uniform and their low diffusion constant can limit the response time of the network. Hence it is necessary to describe the system in space and time.

Techniques have been developed to consider the distribution of the components in space and the stochasticity with which the reactions occur [18, 19, 20]. The reaction-diffusion master equation is the basis for all these techniques, where the space is discretised into a finite number of sub-volumes or voxels. Bimolecular reactions are allowed for two particles in the same voxel, and these particles can move between two voxels by diffusion. In all these techniques work under the assumption that the particles are well stirred in the sub-volume. Hence, these techniques rely on the existence of a time and length scale on which the system is spatially uniform. It is not always clear that there exists such a length and time scale on which the system is well-stirred and well-mixed, especially when the system is deep in the diffusion-limited regime. Indeed, in the past years, this question has been investigated in much detail, and several schemes have been presented that resolve this point by adjusting the voxel size based on the distance between the particles.

This calls for a truly particle-based approach, where the reactants are modelled as ideal (spherical) objects that have a finite size and move by diffusion and the most common algorithm is one based on Brownian dynamics (BD) [21]. BD is a special case of the Langevin equation where the inertial term is zero. In BD, the solute particles are simulated explicitly and the solvent is modelled implicitly. The forces experienced by the particles can be divided into three parts: (i) the drag force, F_{d}, arising due to the motion of the solute particles in a viscous solvent (ii) the potential force, F_{p}, due to interaction of one solute particle with another and (iii) a random force, F_{r}, which due to the stochastic collisions of the solute with the solvent. The fluctuation-dissipation theorem and the Einstein relation, relates , F_{r} to F_{d} both of which depends on the diffusion constant, D, of the solute particles. BD has clear advantages when compared to other grid-based techniques: it naturally allows for anisotropic particles and orientational dynamics and the way in which the solute particles interact *i.e.* the interaction potential, can be easily described. In biochemical networks, as explained earlier, there is a separation of scales due to the low concentration of the reactants. If BD is used to simulate such a system

the algorithm has to use the microscopic time step to propagate particles. Hence the algorithm spends most of the computational effort on propagating particles toward each another before they react. This is one of the main downsides of using BD to simulate biological systems.

To overcome this disadvantage of BD, a new technique, Green's Function Reaction Dynamics(GFRD) was developed [17, 22]. The central idea of GFRD is to decompose the many-body reaction-diffusion problem into sets of one- and two-body problems that can be solved analytically. To this end, single particles and pairs of particles are put into protective domains that do not overlap with each other. For each of the domains, the reaction-diffusion problem is solved analytically using Green's functions. This yields for each domain an event type, which is either a reaction or an exit of a particle from the domain and an event time, which is when this event will happen. These events are put in an event list, which is updated in chronological order. GFRD is thus an event-driven, asynchronous algorithm. Because the domains do not overlap with each other, the stochastic processes of the respective domains are independent, which turns GFRD into an exact algorithm for simulating reaction-diffusion systems at the particle level. It has been shown that at biologically relevant concentrations (micromolar to nanomolar) GFRD performs up to *six* orders of magnitude faster than brute force BD. Despite the incredible computational speed up offered by GFRD, there are disadvantages of using this technique. GFRD assumes idealised particle shape interacting with effective rate constants and neglects the dynamics at the microscopic scale, such as the conformational and orientational dynamics of the particles. At mesoscopic scales where particles are diffusing, the latter can be integrated out without any serious consequences. However, at the microscopic scales it is important to explicitly describe the microscopic dynamics of the particles.

Brute force BD captures the dynamics at the microscopic scales, but wastes most of the computational time propagating the particles toward each another. On the other hand, GFRD particles take huge leaps in space and time, speeding up the diffusion process significantly, but fails to capture the microscopic dynamics. There is not a single technique that can describe the dynamics at both the mesoscopic and the molecular scale which is both accurate and efficient. These problems thus call for a particle-based, multi-scale modelling approach.

1.2. OVERVIEW OF THE THESIS

In *chapter 1*, we propose a novel approach that combines GFRD for simulating the system at the mesocopic scale where particles are far apart, with a microscopic technique such as Langevin Dynamics or Molecular Dynamics, for simulating the system at the microscopic scale where reactants are in close proximity. This scheme defines the regions where (i) the particles are close together and simulated with high microscopic resolution and (ii) the particles are far apart and simulated with lower mesoscopic resolution, adaptively on the fly. The new multi-scale scheme, called Molecular Dynamics Green's Function Reaction Dynamics (MD-GFRD), is generic and can be used to efficiently simulate reaction-diffusion systems at the particle level. MD-GFRD is tested with analytical results for idealised spherical particles interacting via isotropic potentials.

In *chapter 2*, we extend this multiscale MD-GFRD approach to include orientational dynamics, crucial to describe the anisotropic interactions often prevalent in biomolecular systems. We present the novel algorithm focusing on Brownian Dynamics only, although the methodology is generic. We illustrate the novel algorithm using a simple patchy particle model. After validation of the algorithm we discuss its performance. The rotational Brownian Dynamics MD-GFRD multiscale method will open up the possibility for large scale simulations of e.g. protein signalling networks.

In *chapter 3*, we provide microscopic expressions for the intrinsic rate constants for association and dissociation processes of isotropically interacting particles and illustrate how these rates can be computed using rare event simulation techniques. In addition, we address the role of the orientational dynamics for particles interacting via anisotropic potentials.

Using the microscopic expressions derived in chapter 3 , we investigate how the rate of association depends on the strength and the range of the isotropic potential, and the strength of the non-specific attraction in case of the anisotropic potential in *chapter 4*. In addition, we determine the location of the interface where these expressions become valid for anisotropic potentials. In particular, by investigating the orientational distributions of the particles we verify whether the interface at which these distributions become isotropic agrees with the interface predicted by the effective association rate constant. Finally, we discuss how large the intrinsic association rate can become, and what the consequences are for the existence of the diffusion limited regime.

If the substrate has multiple binding sites, a bound enzyme can, besides dissociating into the bulk, also hop to another binding site. However, the expressions derived in chapter 3 are valid only when the substrate has one binding patch. In *chapter 5*, we compute the association, dissociation, and hopping rates as a function of the distance between the binding sites and the rotational diffusion constant. We also determine the effect of blocking of one of these binding sites on the rates. We first derive a new generic expression to calculate transition rates using Forward Flux Sampling, when the states are not necessarily separated by all the interfaces. This is necessary to calculate transition rates between multiple states. Finally, we illustrate this approach by computing these rates for a system in which an enzyme binds to a substrate with two binding sites, and we show how the rates depend on the distance between the binding sites and the orientational diffusion constant of the particles.

Finally, we use the multiscale scheme developed in chapters 1 and 2 along with the rate constants that are calculated in chapter 5 to simulate a biochemical network. As the saying goes 'the proof of pudding is in the eating". To demonstrate the applicability of MD-GFRD, we simulate the response of a MAPK pathway. In each layer of the MAPK cascade, a protein is phosphorylated at two sites. We present a proof-of-principle simulation for a system that is in the linear regime, where the enzyme concentration is much larger than that of the substrate. The aim is to simulate the system under more biologically realistic conditions, in particular more reasonable enzyme and substrate concentrations. We would also like to see the effect of the distance between the binding sites and the orientational diffusion constant on the response of the system.

1

BIBLIOGRAPHY

[1] Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst Biol (Stevenage)* 1:230–236.

[2] Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.* 1:137–151.

[3] Lipková J, Zygalakis KC, Chapman SJ, Erban R (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. *J. Appl. Math.* 71:714–730.

[4] Flegg MB, Chapman SJ, Erban R (2012) The two-regime method for optimizing stochastic reaction-diffusion simulations. *J. R. Soc. Interface* 9:859–868.

[5] Schöneberg J, Noé F (2013) ReaDDy - A Software for Particle-Based Reaction-Diffusion Dynamics in Crowded Cellular Environments. *PLoS ONE* 8:e74261.

[6] Johnson ME, Hummer G (2014) Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. *Phys. Rev. X* 4:031037.

[7] Gillespie DT, Seitaridou E, Gillespie CA (2014) The small-voxel tracking algorithm for simulating chemical reactions among diffusing molecules. *J. Chem. Phys.* 141:234115.

[8] Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *P. Natl. Acad. Sci. USA* 107:2473–2478.

[9] Mugler A, Bailey AG, Takahashi K, Rein Ten Wolde P (2012) Membrane clustering and the role of rebinding in biochemical signaling. *Biophys. J.* 102:1069–1078.

[10] McAdams H, Arkin A (1997) Stochastic mechanisms in gene?expression. *Proceedings of the National Academy of Sciences* 94:814–819.

[11] Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.

[12] van Kampen NG (1992) Stochastic processes in physics and chemistry.

[13] Togashi Y, Kaneko K (2001) Transitions induced by the discreteness of molecules in a small autocatalytic system. *Phys. Rev. Lett.* 86:2459–2462.

[14] Shnerb NM, Louzoun Y, Bettelheim E, Solomon S (2000) The importance of being discrete: Life always wins on the surface. *Proceedings of the National Academy of Sciences* 97:10322–10324.

[15] Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22:403 – 434.

[16] van Zon JS, ten Wolde PR (2005) Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.* 94:128103.

[17] Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a mapk pathway. *Proceedings of the National Academy of Sciences* 107:2473–2478.

[18] Morton-Firth CJ, Bray D (1998) Predicting temporal fluctuations in an intracellular signalling pathway. *Journal of Theoretical Biology* 192:117 – 128.

[19] Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biology* 1:230–236(6).

[20] Baras F, Mansour MM (1997) *Ad. Chem. Phys.* 100:393.

[21] Davidchack R, Ouldridge T, Tretyakov M (2015) New langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* 142.

[22] van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.* 123:234910.

2

COMBINING MOLECULAR DYNAMICS WITH MESOSCOPIC GREEN'S FUNCTION REACTION DYNAMICS SIMULATIONS.

In many reaction-diffusion processes, ranging from biochemical networks, catalysis, to complex self- assembly, the spatial distribution of the reactants and the stochastic character of their interactions are crucial for the macroscopic behavior. The recently developed mesoscopic Green's Function Reaction Dynamics (GFRD) method enables efficient simulation at the particle level provided the microscopic dynamics can be integrated out. Yet, many processes exhibit non-trivial microscopic dynamics that can qualitatively change the macroscopic behavior, calling for an atomistic, microscopic description. We propose a novel approach that combines GFRD for simulating the system at the mesocopic scale where particles are far apart, with a microscopic technique such as Langevin Dynamics or Molecular Dynamics, for simulating the system at the microscopic scale where reactants are in close proximity. This scheme defines the regions where the particles are close together and simulated with high microscopic resolution and those where they are far apart and simulated with lower mesoscopic resolution, adaptively on the fly. The new multi-scale scheme, called Molecular Dynamics Green's Function Reaction Dynamics (MD-GFRD), is generic, and can be used to efficiently simulate reaction-diffusion systems at the particle level.

2.1. INTRODUCTION

Many important reaction-diffusion processes span many orders in length and time scales. Examples are biochemical networks in living cells, catalytic reactions in, e.g. a fuel cell, or dynamics of surfactant/water/oil mixtures. In these systems, the reactants' spatial dynamics at mesoscopic length and time scales of microns and seconds is coupled to the reactions between the molecules at microscopic length and time scales of nanometers and milliseconds. This wide range of length and time scales makes these systems notoriously difficult to simulate.

The conventional strategy to simulate such reaction-diffusion systems, is to coarse-grain the dynamics at molecular scales. One approach is to divide the reaction volume into a number of subvolumes; particles can then react within one volume, but also diffuse from one volume to the next [1]. Another approach is to take into account the particulate nature of the reactants [2, 3, 4, 5, 6, 7]. In these particle-based schemes, coarse graining typically involves describing the reactants and their interactions by effective cross sections, diffusion constants and rate constants. This approach drastically coarse grains the dynamics at molecular scales. However, it is now clear that the dynamics at the molecular scale can qualitatively change the macroscopic behavior of the system. It has been shown that the rapid rebinding between a catalyst and its substrate can drastically change the rate of product formation[8], especially when one or both reactants are on a surface[9]. In such cases, the binding between substrate and enzyme, the conformational dynamics of the molecules, the formation and release of the product, the orientational diffusion of the reactants and products has to be described explicitly, and cannot be integrated out. Another approach, to simulate such systems is using (ab-initio) molecular dynamics at all scales. This, while possible in principle, would be highly inefficient if not impossible in practice, since much CPU time would be spent on propagating the reactants towards one another before they can react. There is not a single technique, that can describe the dynamics at both the system and the molecular scale which is both accurate and efficient. These problems thus call for a particle-based, multi-scale modeling approach.

In this article we propose such a particle-based, multi-scale scheme. When far apart, the particles are propagated toward each other using a mesoscopic technique. Here, the dynamics at the molecular scales are not important since the particles are far apart. When the particles are close to each other, the algorithm switches to a microscopic/atomistic description of the particle dynamics, which takes into account the anisotropy of the particles and their conformational dynamics. In the multi-scale scheme proposed here, the particle dynamics at the different scales are thus described by different techniques, and the algorithm switches seamlessly between them based on predefined scenarios. The algorithm can be applied to any reaction-diffusion system, but it becomes particularly efficient when there is a clear separation in length and time scales between the dynamics at the molecular and system level.

The challenges to develop a particle-based multi-scale scheme are three-fold. First, we need a mesoscopic description of particle dynamics when the particles are far apart, to simulate particle diffusion. Second, we need a microscopic description of particle dynamics when the particles come close together, to simulate reactions. Third, we need to couple the two techniques to obtain a multi-scale scheme. While the scheme presented

here can be applied to any particle-based reaction-diffusion model, we describe it here with the application to cellular biochemical networks in mind.

In cellular systems, the mesoscopic particle diffusion typically occurs on time scales of seconds and length scales of microns. A particle based algorithm, based on brute-force Brownian Dynamics[10, 3, 4, 5], would be an obvious choice to describe particle diffusion at the cellular scale. In this approach, the particles have a certain (idealized) size and shape, move with a prescribed diffusion constant, and can react with each other upon contact. This approach, while much more efficient than straightforward Molecular Dynamics, is still highly inefficient, because in these cellular systems the concentrations of the reactants are often low, *i.e.* in the nM or μM range, which means that most CPU time is spent on propagating the particles towards one another. To overcome this problem, Green's Function Reaction Dynamics (GFRD) was developed [11, 8]. The crux of GFRD is to decompose the many-body reaction-diffusion problem into sets of one- and two-body problems that can be solved analytically using Green's Functions. These Green's Functions are then used to set up an exact, event-driven algorithm, in which the particles can make large jumps in time and space when they are far apart from each other. Indeed, GFRD can be up to 6 orders of magnitude faster than Brownian Dynamics under biologically relevant conditions. GFRD assumes idealized particle shape and neglects the dynamics at the microscopic scale, such as the conformational and orientational dynamics of the particles. However, at the mesoscopic scale the latter can be integrated out. In our multi-scale scheme we therefore use GFRD to simulate mesoscopic particle diffusion. A detailed description of GFRD is given in the next section.

The microscopic particle dynamics typically occurs on time scales of milliseconds and length scales of nanometers, given by the typical size of proteins. When the particles come close to each other, their encounters can be simulated using microscopic techniques such as Molecular Dynamics, Brownian dynamics or an accurate coarse grained representation of the microscopic dynamics such as Markov State Modeling(MSM) [12, 13]. In the scheme presented here, we use Brownian Dynamics (BD) to treat the dynamics at the molecular scale. BD is Langevin dynamics in the overdamped regime, which is a stochastic dynamics scheme that describes the solutes explicitly, but treats the solvent molecules only implicitly via the stochastic and friction forces that they exert on the solutes. In this article, the second order quasi symplectic integrator suggested by[14] is used to numerically solve the BD equations. This integrator can propagate the orientational dynamics of the particles, although here we limit ourselves to simple Lennard-Jones particles with isotropic interactions.

In the multi-scale Molecular Dynamics - Green's Function Reaction Dynamics (MD-GFRD) scheme that we propose here, we use BD to simulate the reactants that are within a reaction distance, and we use GFRD to propagate the particles that are beyond this distance. MD-GFRD thus exploits the power of GFRD to allow particles to make large jumps in time and space when they are far away from each other. Importantly, in contrast to other multi-scale schemes [15, 16], the regions where the particles are close together and simulated with high microscopic resolution, and the parts of space where they are far apart and simulated with lower mesososcopic resolution, are defined adaptively on the fly. This makes MD-GFRD, under biologically relevant conditions, orders of magnitude more efficient than brute-force BD with the same microscopic resolution. While

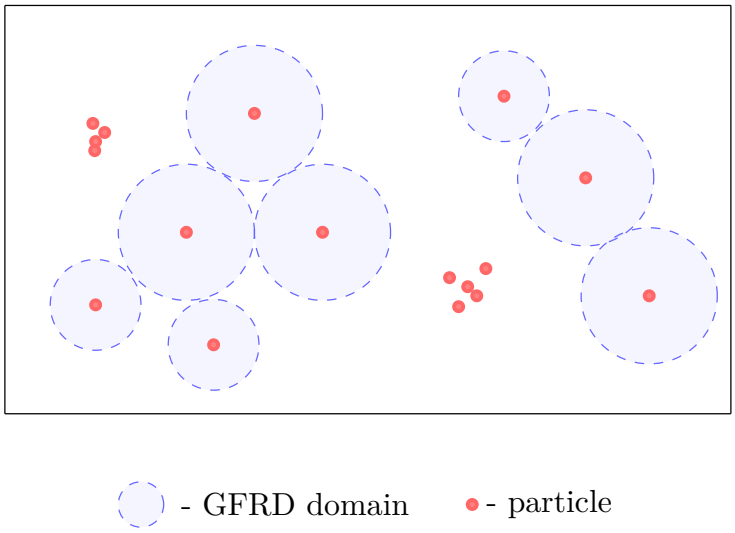


Figure 2.1: The Multi-scale scheme: particles that are far away from each other are surrounded by an GFRD domain, and particles close to each other are in the BD regime and propagated at the same time.

we use BD for simulating the particles at close distances, the principal idea is generic. In the scheme presented here, GFRD can also be combined with other schemes such as atomistic MD or MSM.

2.2. THE MULTI-SCALE SCHEME

2.2.1. OVERVIEW

Before we introduce MD-GFRD, it is instructive to first briefly describe GFRD [8]. The central idea of GFRD is to decompose the many-body reaction-diffusion problem into sets of one- and two-body problems that can be solved analytically [11, 17]. To this end, single particles and pairs of particles are put into protective domains that do not overlap with each other [8, 18]. For each of the domains, the reaction-diffusion problem is solved analytically using Green's functions. This yields for each domain an event *type*, which is either a reaction or an exit of a particle from the domain, and an event *time*, which is when this event will happen. These events are put in an event list, which is updated in chronological order. GFRD is thus an event-driven, asynchronous algorithm. Because the domains do not overlap with each other, the stochastic processes of the respective domains are independent, which turns GFRD into an exact algorithm for simulating reaction-diffusion systems at the particle level.

In the scheme proposed here, we only keep the Single domains of the original GFRD scheme. Single particles that are sufficiently far away from other particles according to a predefined cut-off distance are put into protective domains, as shown in Fig. 3.1. For

each of these domains, the algorithm determines, as in the conventional GFRD scheme, the next-event type, which is either a mono-molecular decay reaction or an exit of the particle from the domain, and the corresponding next-event time. Each domain is then updated only when its time has come, *i.e.* when the simulation time has reached the event time of the domain. This allows MD-GFRD to make large jumps in time and space when the domains are large. It is the origin of the high efficiency of the scheme.

The other particles, which are within the predefined cut-off distance from another particle or GFRD domain, are propagated with conventional BD. They are not put into protective domains and, as in conventional BD, they are propagated collectively using small time steps. The particles interact with each other via an interaction-potential, allowing for a detailed microscopic description of the conformational and orientational dynamics of the particles that are in close vicinity of each other. This dynamics also implements the bimolecular association reactions, which in the original GFRD scheme were treated via the Pair domains.

MD-GFRD as presented here can thus incorporate two types of reactions: mono-molecular decay reactions and bi-molecular association reactions. How these reactions are simulated is described in detail in the sections below. Here we give a brief overview.

In a bi-molecular reaction two particles A and B can react with an intrinsic rate constant k_a to form products:

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_a} \mathrm{C} + \mathrm{D} + \ldots \tag{2.1}$$

Only particles in the BD regime can undergo a bi-molecular reaction. Bi-molecular reactions are discussed in detail in Sec. 2.2.3.

In a mono-molecular reaction the particle can instantaneously decay into products with an intrinsic rate constant k_d according to:

$$\mathrm{A} \xrightarrow{k_d} \mathrm{B} + \mathrm{C} + \ldots \tag{2.2}$$

The decay reaction is assumed to be a Poisson process. A particle in either the GFRD or the BD regime can decay into its products.

By combining Eqs. 2.1 and 2.2 essentially any complex biochemical network can be simulated. Here we present MD-GFRD for the simple case of three species A, B and C, which react according to

$$\mathrm{A} + \mathrm{B} \underset{k_\mathrm{d}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{C}.$$

Two particles A and B are in the bound state if they are close enough together such that the inter-particle force between them is strong enough to create the dimer A-B. When the monomers are beyond the interaction range of the potential, the inter-particle force no longer holds the dimer A-B together, and the particles enter the unbound state. In MD-GFRD, the association reaction is always simulated with BD. Also the dissociation reaction could be simulated with BD: we could explicitly simulate the monomers in the dimer A-B, until they dissociate into A and B. However, when the dissociation time is longer than the time it takes for the dimer to randomize its orientation, we can replace the dimer with a single particle C, which can then be simulated with GFRD. This reduces the time to simulate the dimer, making the scheme more efficient. The C particle can

then decay with a rate constant k_d, yielding the monomers A and B again. The value of k_d can be obtained by a preliminary simulation or analytically.

At the heart of MD-GFRD, there is thus an event list, ordered based on the next event times of the particles. These next events can be: (1) Escape of a particle from a GFRD domain, (2) Decay of a particle in a GFRD domain, (3) Decay of a Brownian particle. The scheme then proceeds as follows. The particles in the BD regime are propagated. During this time, attempts are continually being made to put BD particles in the GFRD domain. If an attempt is successful, the event list is updated. This dynamics continues until the simulation time reaches the time of the first event in the event list. This event is then executed, the system is updated accordingly, for the updated particles the next event types and times are determined and inserted into the event list, and the BD particles are propagated again, and so on. The scheme becomes particularly powerful when most particles are in GFRD domains. A key objective is thus to keep the number of BD particles to a minimum. In the subsequent section, we describe in more detail the different steps of the algorithm.

2.2.2. GREEN'S FUNCTION REACTION DYNAMICS(GFRD)

A particle in a GFRD domain has, as discussed in the previous section, two possible next events. It can either exit the domain or decay into other species (C$\xrightarrow{k_d}$A+B). For each of these tentative events, we compute the tentative event time. The next-event type for the domain is then the tentative event with the smallest tentative event time; this time then also becomes the next-event time for that domain. We now briefly describe how we compute the event times for the two event types.

ESCAPE FROM THE DOMAIN

The Einstein diffusion equation describes the motion of a freely diffusing spherical particle:

$$\partial_t p(\mathbf{r}, t|\mathbf{r}_0, t_0) = D\nabla^2 p(\mathbf{r}, t|\mathbf{r}_0, t_0), \tag{2.3}$$

where $p(\mathbf{r}, t|\mathbf{r}_0, t_0)$ denotes the probability that a freely diffusing spherical particle will be located at $\mathbf{r}$ at time t, given that it was at $\mathbf{r}_0$ at time t_0. We obtain the solution to Eq.2.3, by imposing the following initial and boundary conditions:

$$p(\mathbf{r}, t_0|\mathbf{r}_0, t_0) = \delta(r - r_0), \tag{2.4a}$$

$$p(|\mathbf{r} - \mathbf{r}_0| = a, t_0|\mathbf{r}_0, t_0) = 0, \tag{2.4b}$$

where a is the radius of the domain $\mathscr{D}$ and δ is the Dirac delta function. The solution obtained is the Green's function $p(\mathbf{r}, t|\mathbf{r}_0, t_0)$, and integrating this over the whole domain gives the survival probability $S(t)$, which is the probability that the particle is still in the domain at time t:

$$S(t) = \int_{\mathscr{D}} p(\mathbf{r}, t|\mathbf{r}_0, t_0)\, d\mathbf{r}. \tag{2.5}$$

The propensity function, which is the probability per unit amount of time that the particle escapes the domain for the first time, $q_e(t)$, is related to the survival probability by

$$q_e(t) = \frac{-\partial S(t)}{\partial t}. \tag{2.6}$$

The cumulative distribution of $q(t)$ is given by

$$Q(t) = \int_{t_0}^{t} q(t)dt = 1 - S(t), \tag{2.7}$$

which can be used to sample next escape times. The tentative next escape time, τ_e, is sampled from $Q(t)$ by numerically solving,

$$Q(\tau_e) = 1 - S(\tau_e) = \mathscr{R}_e, \tag{2.8}$$

where $\mathscr{R}_e \in [0,1]$ is a uniformly distributed random number.

MONO-MOLECULAR DECAY

The propensity function for the next dissociation reaction, is exponentially distributed and is given by,

$$q_d(t) = k_d e^{-k_d t}, \tag{2.9}$$

where k_d is the rate of the dissociation reaction. The tentative next event time for the dissociation reaction, τ_d, is sampled using

$$\tau_d = -k_d \ln(\mathscr{R}_d), \tag{2.10}$$

where $\mathscr{R}_d \in [0,1]$ is a uniformly distributed random number.

NEXT EVENT TIME

The next event time, τ, is the time at which the particle either dissociates or escapes from the domain, given that the particle is in the center of the domain. As soon as a domain is constructed, τ_d and τ_e are evaluated using Eq. 7.9 and Eq. 2.8 respectively. The next event time is then calculated as:

$$\tau = \min(\tau_d, \tau_e). \tag{2.11}$$

DOMAIN BURSTING

In the course of the simulation, a particle that has just escaped from the GFRD domain may have come close to another GFRD domain. In principle a small domain could be constructed around this particle, but this would make the scheme inefficient. Moreover, a particle in the BD regime, which is propagated with a fixed finite time step, may penetrate a GFRD domain to which it is close. In both cases, the GFRD domain of interest—the domain to which the particle is close—is *burst* to generate more space. More specifically, a GFRD domain is burst when: (1) A particle that has just escaped from a GFRD domain is within a distance r_{BD} (see Fig. 2.4) from the boundary of the domain of interest; (2) A particle in the Brownian regime comes with in a distance r_{BD} from the boundary of the domain of interest; (3) A particle in a domain decays into its products.

Bursting means generating a position of the particle at time τ_b before it escapes from the domain, where $\tau_b < \tau_e$. The radial position is sampled by numerically solving

$$P_{S(\tau_b)}(r_b) = \mathscr{R}_b, \tag{2.12}$$

where $\mathscr{R}_b \in [0,1]$ is a uniformly distributed random number and the cumulative conditional PDF

$$P_{S(\tau_b)}(r_b) = \frac{1}{S(\tau_b)} \int_0^{r_b} \int_0^{2\pi} \int_0^{\pi} p(r,\theta,\phi,\tau_b|\mathbf{r}_0) r^2 \sin(\theta) d\theta d\phi dr. \tag{2.13}$$

$P_{S(\tau_b)}(r_b)$ is normalized with the survival probability $S(\tau_b)$, since we know that the particle is inside the domain at time τ_b.

2.2.3. BROWNIAN DYNAMICS

Particle dynamics at the molecular scale is generated by numerically propagating the BD equations. BD is the overdamped version of Langevin Dynamics, meaning that we propagate only the positions of the particles, and not their velocities. A second order quasi symplectic integrator is used to numerically integrate the BD equations [14]. This integrator is particularly well suited for taking into account the orientational dynamics, although we will not employ this feature here. For a detailed description of the algorithm please see Ref.[14].

THE LENNARD-JONES POTENTIAL

The particles in the BD regime influence each other through a potential. To present the algorithm, we use simple particles that interact via a conventional, isotropic Lennard-Jones potential. In future work we will describe how anisotropic interactions, including orientational dynamics, can be included into MD-GFRD. The Lennard-Jones potential, $V(r)$, shown in Fig. 2.2, is given by

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.14}$$

where r is the inter-particle distance.

The potential has two parameters, σ and ϵ. The parameter σ is the inter-particle cross section, *i.e.* the distance r at which the potential, $V(r)$, is zero. The parameter ϵ is the potential well depth, which is a measure of the interaction strength between the particles.

For all other interactions, A-A, B-B, C-C, C-A, C-B, we use the following repulsive potential

$$V_{WCA}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon \text{ if } r < 2^{\frac{1}{6}}\sigma \tag{2.15}$$

FREE ENERGY OF TWO INTERACTING PARTICLES

BD simulates the reaction between the two particles. To achieve this we need a definition of the bound and the unbound states. To this end, we will use the free energy (F) of a two-particle system as a function of its distance. The probability density that two particles interacting via a potential $V(r)$ are at a distance r is given by

$$P(r) = \frac{1}{Z} r^2 e^{-\frac{V(r)}{k_B T}}, \tag{2.16}$$

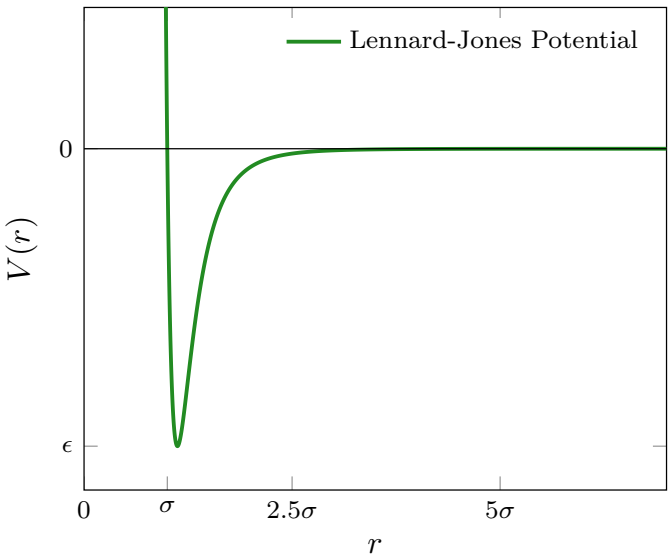


Figure 2.2: Lennard-Jones Potential as given by Eq. 2.14. The parameters used for our simulations are: $\epsilon = 10k_BT$ and $\sigma = 5$nm

where k_B is the Boltzmann constant, T is the temperature and Z the partition function given by

$$Z = \int r^2 e^{-\frac{V(r)}{k_B T}} dr. \tag{2.17}$$

The Landau free energy, $F(r)$, of the system as a function of r is given by the negative log of the $P(r)$

$$F(r) = -k_B T \ln[P(r)] = -2k_B T \ln(r) + V(r) + k_B T \ln Z. \tag{2.18}$$

We note that the free energy contains a log(r) term due to the Jacobian in three dimensions and will use this free energy $F(r)$ to define the bound and the unbound state, as described in the following section.

REACTIONS AND FIRST-ESCAPE

The scheme handles reactions of the type

$$\mathrm{A} + \mathrm{B} \underset{k_\mathrm{d}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{C} \tag{2.19}$$

where k_a and k_d are the association and dissociation rates, respectively. The free energy landscape, Fig. 2.3 is used to determine whether the particles are bound or not. The free energy landscape shows a peak for an inter-particle distance r_p. If the inter-particle distance is less than r_p, the particles are in the bound state, while if it is larger, the particles are in the unbound state.

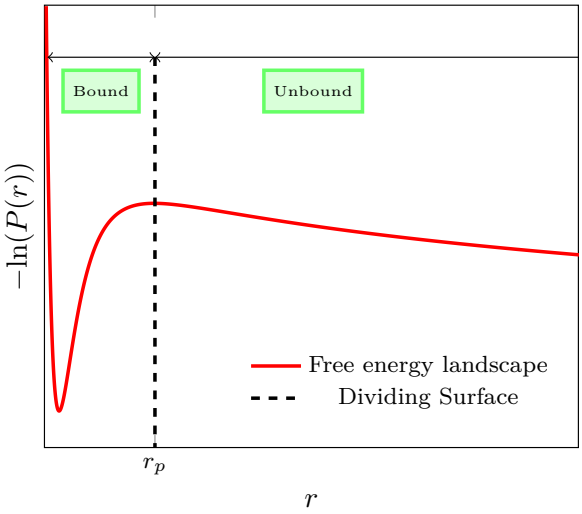


Figure 2.3: Free Energy landscape showing a peak at r_p. This forms a barrier between the bound and the unbound state. Reactions occur based on the inter-particle distance, r. If $r < r_p$, then the particles associate, and are in the bound state. If $r > r_p$, then the particles dissociate, and go to the unbound state. The time taken for a particle starting at the bottom of the potential well, to cross the barrier, is called as the dissociation time.

The central idea is that the association reactions are implemented via BD. As described in detail in the next section, when two particles come within a predefined cut-off distance, they are simulated with BD. An association reaction between the two particles then occurs when these two particles spontaneously cross the free-energy barrier at r_p.

In principle, the algorithm could continue to propagate these two particles that are now in the bound state, with BD. However, when the interaction between the two particles is strong, as is typically the case in cellular systems, then it will take a relatively long time before they cross the free-energy barrier and dissociate again. During this time, the two particles will simply rattle around each other in the potential well, while their center-of-mass slowly diffuses. If the dissociation time is longer than the time it takes for the dimer to reorient itself, then it becomes advantageous to treat the particles as a single particle C, centered at the center-of-mass of the particles A and B. This C particle diffuses with a diffusion constant, which depends on the nature of the interaction potential; if the interaction potential depends on the inter-particle distance only, the diffusion constant is simply the sum of the diffusion constants of particles A and B, respectively [8]. Around this particle C, a GFRD domain can be built if the space is available, which would significantly speed up the simulations. The C particle then diffuses, either explicitly with BD or implicitly with GFRD, until it dissociates again into the monomers A and B. Upon dissociation, the monomers are put at a close distance as described below, and in a random orientation (if necessary), which is justified when the dissociation time is longer than the reorientation time of the dimer.

When the dimer A-B is treated as a single particle C, we need the time τ_{d} at which it dissociates into the monomers A and B again. When the potential well is deep enough, the distribution of dissociation times τ_{d} will be exponential and given by Eq.3.5. Clearly, we can generate the next dissociation time τ_d using Eq. 7.9, once we know the dissociation rate k_{d}. To determine k_{d}, we perform prior to the MD-GFRD simulations, a brute-force BD simulation of two particles, in which we compute the distribution of dissociation times. The rate constant k_{d} is then obtained by fitting the measured distribution to Eq. 3.5.

Lastly, while r_p defines the boundary between the bound and the unbound state, we do not use this dividing surface to switch between the bound and unbound state during the simulations. The reason is that with a single dividing surface (located at r_p) the system would rapidly recross this dividing surface many times before the two particles either escape from each other or settle in the potential well. Inspired by the indicator functions used in rare-event simulation techniques [19, 20, 21, 22], we therefore define two dividing surfaces B1 and B2, as shown in Fig. 2.4. Two unbound particles need to cross the dividing surface B1 to enter the bound state. But once bound, the (monomers of) the dimer has to cross the dividing surface B2 for dissociation. By putting B1 and B2 a few $k_B T$s below the top of the barrier, the rapid re-bindings are eliminated. And since the time the system spends near the top of the barrier is small compared to that in the free energy minima, this procedure does not affect the value of the rate constants k_{a} and k_{d} [19].

2.2.4. COUPLING THE TWO SCHEMES

To couple BD and GFRD we need to address two questions: When should a BD particle be put into a protective GFRD domain? And when should a particle that escapes from a GFRD domain be treated as a BD particle? To start with the first question, if during the BD integration, a BD particle has moved away from the other BD particles and GFRD domains farther than a distance r_{GFRD} , then a GFRD domain will be constructed around the particle.

When should the scheme switch back from GFRD to BD? A particle in a GFRD domain has two possible next event types (Fig. 2.5). It can either dissociate or escape from the domain. If the particle C in a GFRD domain dissociates, it is replaced by particles A and B. The particles A and B are now placed at a distance r_{d} (Fig. 2.4) and propagated with BD. When the particle escapes from the GFRD domain, the position of the particle is updated to a random position on the boundary of the domain and the domain is removed. We then check if a new domain can be constructed on the particle that escaped. To this end we measure the distance from the particle of interest to its nearest neighbor, to determine if there is enough space to construct a domain. The nearest neighbor to the particle of interest can either be a BD particle or a GFRD domain. If the nearest neighbor is a BD particle, and the distance r from the BD particle to the particle of interest is greater than $2r_{\mathrm{BD}}$ (Fig. 2.4), we construct a GFRD domain of radius $0.5r$ around the particle of interest. If r is less than $2r_{\mathrm{BD}}$ then we propagate the particle of interest with BD since there is no space to build a domain. On the other hand, if the nearest neighbor is a GFRD domain, and the distance r from the particle of interest to the boundary of the GFRD domain is greater than r_{BD}, we construct a domain of radius r around the par-

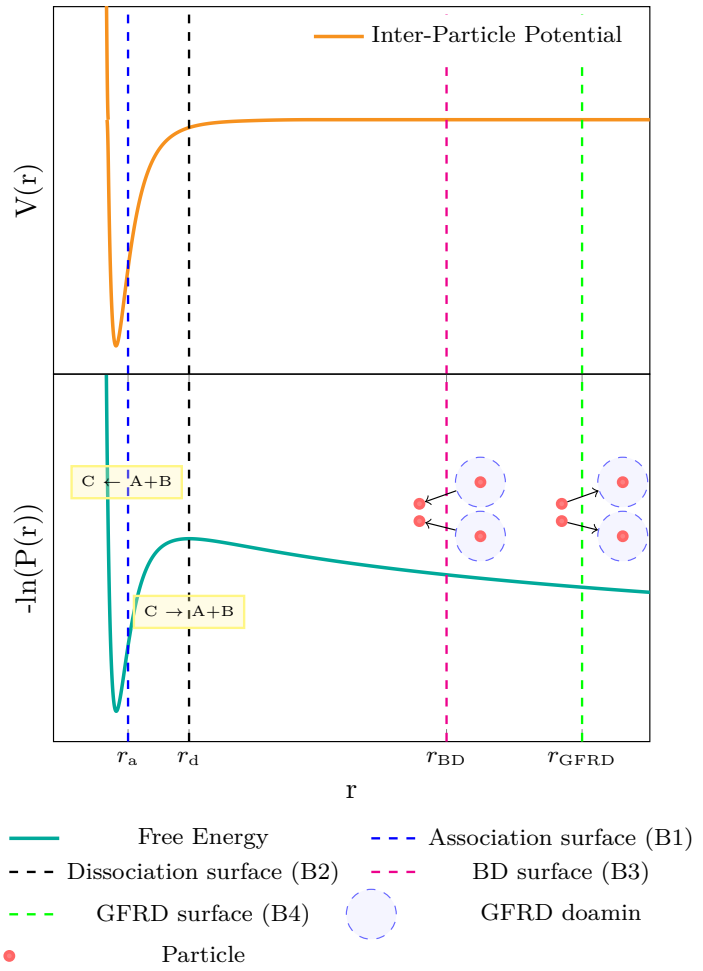


Figure 2.4: Dividing surfaces and their functions. If two particles in the BD regime come close to each other, such that $r < r_a$, then the particles associate. The two particles are replaced by a single particle of the appropriate species. The next dissociation time is drawn from the distribution given by Eq. 3.5. When a particle dissociates, we replace the single particle by its products and place them at a distance $r = r_d$. When two BD particles move far from each other, such that $r > r_{GFRD}$, we build domains over the particles and put them in the GFRD regime. When two particles in the GFRD regime come close, such that $r < r_{BD}$, we put the particles in the BD regime. See Fig.2.5 for more details.

ticle of interest. If however r is less than r_{BD}, the particle of interest bursts the domain, as discussed in 2.2.2. When a domain is burst, a new position is generated for the particle inside the domain and the particle is converted to a BD particle. It could happen that this newly created BD particle is within a distance less than r_{BD} from a neighboring domain. In such a case, the newly created BD particle bursts the neighboring domain. This can lead to a cascade of domain bursting until finally no BD particle is within a distance of r_{BD} from any domain (Fig. 2.5, pop-up panel). Now we loop over all the newly created BD particles and attempt to build domains around them. For each newly created BD particle, we find the distance to its nearest neighbor. If the nearest neighbor is a GFRD domain, then we build a domain of radius r equal to the distance to the boundary of that domain, since we know that this distance r is larger than r_{BD} (otherwise the domain would have been burst). If the nearest neighbor is instead a BD particle, then, if the distance r is greater than $2r_{BD}$, a domain of radius r is constructed, else the particle is propagated with BD.

As sketched in Fig. 2.4, r_{GFRD} is taken to be larger than r_{BD}. This is important, because it prevents the system from rapidly switching between BD and GFRD many times before firmly settling in either the BD or GFRD mode.

2.2.5. OUTLINE OF THE ALGORITHM

This section gives an outline of the multi-scale scheme. In the simulations there are two types of particles: BD particles and GFRD particles and three types of species; A, B and C. A BD particle has only one type of next event namely dissociation, whereas a GFRD particle's next event type maybe a dissociation or an escape from its domain. The algorithm is described in Alg. 2.

Algorithm 1 The heart of the Multi-scale scheme. $\mathscr{P}_n$ is the list of all particles, τ_n the next event times, $\mathscr{T}_n$ the next event types, t_{sim} the time passed since the start of the simulation, δt the time step interval used for the BD integrator and r the inter-particle distance. When a domain is burst or when a particle escapes from a domain, the particle might burst a neighboring domain and this may continue. This is called successive bursting. create_domains is a user defined routine which deals with successive bursting and attempts to build domains around particles. Fig. 2.5 outlines the switching from GFRD to BD, including a description of the function create_domains.

Initialization:
$t_{sim} \leftarrow 0$
$\mathscr{P}_n \leftarrow$ read in initial particle positions and velocities
create_domains for all BD particles p_l in $\mathscr{P}_n$
rearrange $\mathscr{P}_n$ based on τ_n

Main loop:

while $t_{sim} < t_{end}$ **do**

BD section:

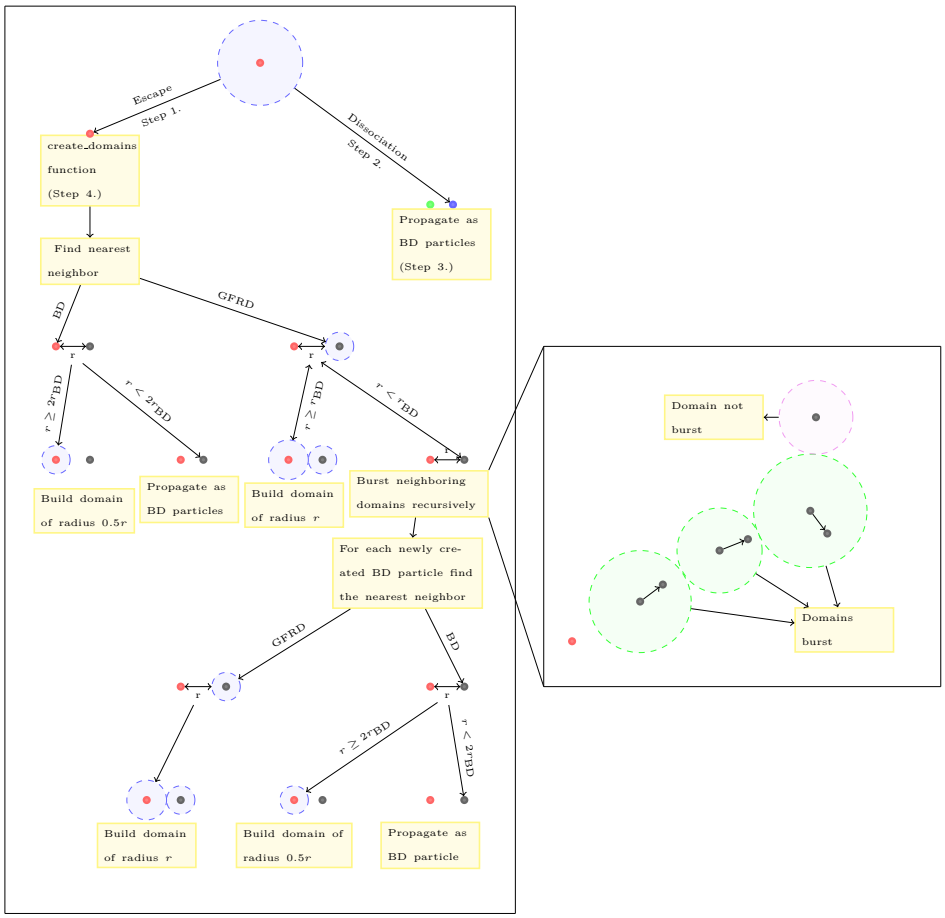


Figure 2.5: The switching from GFRD to BD. A particle in a GFRD domain has two possible next event types. It can either dissociate or escape from the domain. If the particle dissociates, the products are propagated with BD since they are placed at a distance $\mathscr{D}_d < 2r_{\mathrm{BD}}$. If the particle escapes from the domain, the nearest neighbor to this particle of interest is determined. The nearest neighbor can either be a BD particle or a GFRD domain.

(continued next page)

Figure 2.5: (*continuing from previous page*) If the nearest neighbor is a BD particle and the distance r from this particle to the particle of interest is greater than $2r_{\mathrm{BD}}$ a domain of radius $0.5r$ is built around the particle of interest and if $r < 2r_{\mathrm{BD}}$ both particles are propagated with BD. On the other hand, if the nearest neighbor is a GFRD domain, the distance r from the particle of interest to the boundary of the GFRD domain is measured. If $r \geq r_{\mathrm{BD}}$, a domain of radius r is built around the particle of interest. If however $r < r_{\mathrm{BD}}$, the GFRD domain is burst. Bursting a domain leads to a BD particle with a newly sampled position. Now this newly created BD particle due to the burst domain, can in turn burst another neighboring domain leading to another BD particle and this can happen recursively until finally no BD particle is within a distance of r_{BD} from any domain. This is shown in the second panel of the figure. After all possible domains are burst, for each of the newly created BD particles we find the nearest neighbor. If for a BD particle the nearest neighbor is a GFRD domain, we build a domain of radius r around the BD particle since we know that they are separated by at least r_{BD}. If for a BD particle, the nearest neighbor is another BD particle, the distance r between the two particles is measured. If $r \geq 2r_{\mathrm{BD}}$ a domain of radius $0.5r$ is built around the BD particle of interest, else if $r < 2r_{\mathrm{BD}}$ the BD particle of interest is propagated with BD. Please note that some of the steps are numbered, to facilitate the description of the algorithm in the text; the numbers do not correspond to any particular order in which the steps need to be executed.

---

if number of BD particles > 0 **then**

Check if any BD particle dissociates:
for all BD C particles p_l in $\mathscr{P}_n$ **do**
if $\mathscr{T}_l$==dissociation and $t_{sim} \geq \tau_l$ **then**
dissociate p_l and update $\mathscr{P}_n$
place the products at r_d(see Fig. 2.4)
end if
end for

BD numerical integrator:
for all BD particles p_l in $\mathscr{P}_n$ **do**
calculate forces
update position
end for
$t_{sim} += \delta t$

Check if any two BD particles react:
for all BD A and B particles p_l in $\mathscr{P}_n$ **do**
for all BD A and B particles p_m in $\mathscr{P}_n - p_l$ **do**
if $r \leq r_a$(see Fig. 2.4) **then**
replace A and B (if possible) by C
update $\mathscr{P}_n$
end if
end for
end for

```
    create_domains for all BD particles p_l in P_n
    rearrange P_n based on τ_n
  end if

  GFRD section:
  if number of GFRD particles > 0 then
    for p_0, the first particle in the P_n
    if t_sim ≥ τ_0 or number of BD particles == 0 then

      if T_0 == escape (Step 1. in Fig. 2.5) then
        t_sim ← τ_0
        update position of p_0
        make p_0 an BD particle
        create_domains for all BD particles p_l in P_n (Step 4. in Fig. 2.5)
        rearrange P_n based on τ_n
      end if

      if T_0 == dissociation (Step 2. in Fig. 2.5) then
        burst the domain
        place the products at r_d (see Fig. 2.4; Step 3. in Fig. 2.5)
        rearrange P_n based on τ_n
      end if

    end if
  end if
end while
```

2.3. RESULTS AND DISCUSSION

In this section we discuss the tests performed to check the validity of the multi-scale scheme. Tests are performed for the system

$$\mathrm{A} + \mathrm{B} \underset{k_\mathrm{d}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{C}. \tag{2.20}$$

First we discuss how we determine the rate constant, k_d. Then using this value of k_d, we perform a multi-scale MD-GFRD simulation where the probability of the two species, A and B, being bound is calculated for two cases:(1) One A particle surrounded by many B particles; (2) Many A particles surrounded by many B particles. In both cases the simulation results are validated with analytical results. Next, the power spectrum is evaluated for the multi-scale simulation and is compared to that of a brute force BD simulation. Finally, the performance of the multi-scale simulation is compared to that of a brute force BD simulation. For all MD-GFRD simulations the box width is 120σ, mass=5kDa, $r_{\mathrm{BD}} = 3\sigma$, $r_{\mathrm{GFRD}} = 5\sigma$ and time step for the BD integrator δt =10ns. The diffusion constant

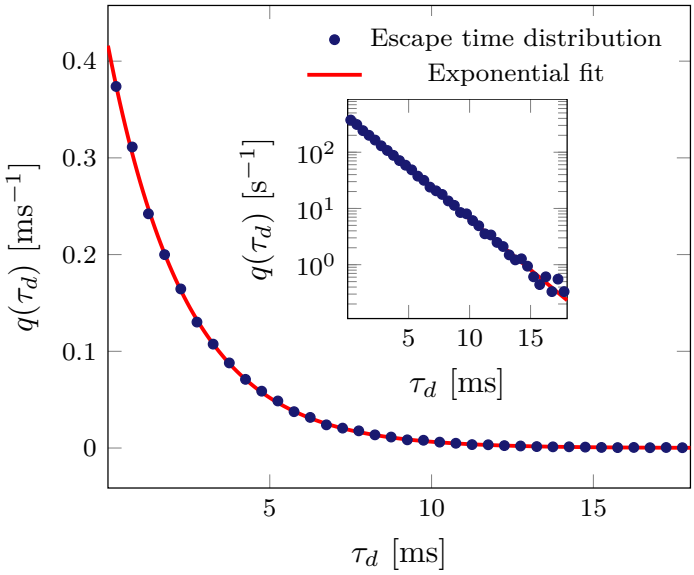


Figure 2.6: Escape Time distribution fitted with an exponential curve. From the fit we obtain k_d=0.42ms^{-1}. $q(\tau_d)$ is the probability per unit time, that the particle dissociates at time τ_d. The simulation was performed for a well depth of $\epsilon = 10 k_B T$. The inset shows a semi-log plot of the exponential fit.

D, the number of particles N and the interaction strength ϵ are varied and described below.

2.3.1. RATE CONSTANT

When the dimer A-B is treated as a C particle, then, as discussed in Sec. 2.2.3, we need the value of the rate constant, k_d. For this purpose, a brute force BD preliminary simulation with two particles in a box, with periodic boundaries, is performed. The particles are in a bound state if the inter-particle distance $r < r_a$, see Fig. 2.4. If the particles move apart and $r > r_d$, then the particles are unbound. The time, called the escape time, τ_{es}, taken for a bound particle to become unbound is recorded. A typical histogram of the escape times is shown in Fig. 2.6. If the well of the inter-particle potential is deep enough, the escape times will follow an exponential distribution Eq. 3.5. The histogram of the escape times is fitted to an exponential and the value of k_d is calculated.

2.3.2. BIMOLECULAR REACTIONS

As a first test of the multi-scale scheme, we simulate a reversible bimolecular reaction, Eq. 2.20. In these simulations, we start of with two species of particles, A and B. One A and one B can react to form a dimer. The dissociation time of the dimer is typically much larger than the time for the dimer A-B to randomize its orientation, so that the A-B dimer

can be treated as a single particle C around which a GFRD domain can be built. Also, one C particle can dissociate to form one A and one B. Here we assume that particles of similar species do not interact with each other and species C does not interact with any species. We test the scheme for two different scenarios as described below.

ONE A MANY B

In the first case, one particle of species A, is surrounded by N_B particles of species B, in a box of volume V, with periodic boundary conditions. This means that the number of C particles, N_C, is either zero or one and that at any moment in time we will have $1-N_C$ A particles and N_B-N_C B particles. The simulation calculates the average time that the particle A is bound. From this, we calculate the probability $\mathscr{P}_b$ of the particle A being bound to a particle B. We repeat the simulation for different N_B keeping V constant. We validate the value of $\mathscr{P}_b$ obtained from the simulation to the analytical value given by,

$$\mathscr{P}_b = \frac{1}{1+\dfrac{K_D V}{N_B}} \tag{2.21}$$

where K_D is the dissociation constant given by,

$$K_D = \frac{1}{4\pi \int_0^{r_p} r^2 e^{-\beta V(r)} dr} \tag{2.22}$$

where r_p is defined by Fig. 2.3 and $V(r)$ is given by Eq. 2.14. Fig. 2.7 shows excellent agreement between the simulation and analytical results, for two different values of inter-particle potential strength, ϵ (see Eq. 2.14).

MANY A MANY B

The second case is a slightly more complex system, where $N_A > 1$ particles of species A and N_B particles of species B are simulated in a box, with periodic boundary conditions. This means that, at any moment we will have $N_A - N_C$ A particles, $N_B - N_C$ B particles and N_C C particles. In this case the the maximum number of C particles is $\min(N_A, N_B)$, and the probability $\mathscr{P}_b$ of particle A being bound is given by,

$$\mathscr{P}_b = \frac{\langle N_C \rangle}{N_A}, \tag{2.23}$$

where $\langle N_C \rangle$ is the average number of C particles. Using mean field chemical rate equations we have,

$$\langle N_C \rangle = \frac{M - \sqrt{M^2 - 4N_A N_B}}{2},$$
$$M = (N_A + N_B) + K_D V, \tag{2.24}$$

where K_D is given by Eq. 2.22

Fig. 2.8 shows that the simulation results of MD-GFRD agree very well with the analytical predictions of Eq. 2.24. The top panel shows that the MD-GFRD simulation results also agree with those of brute-force BD. These results show that MD-GFRD successfully predicts average quantities.

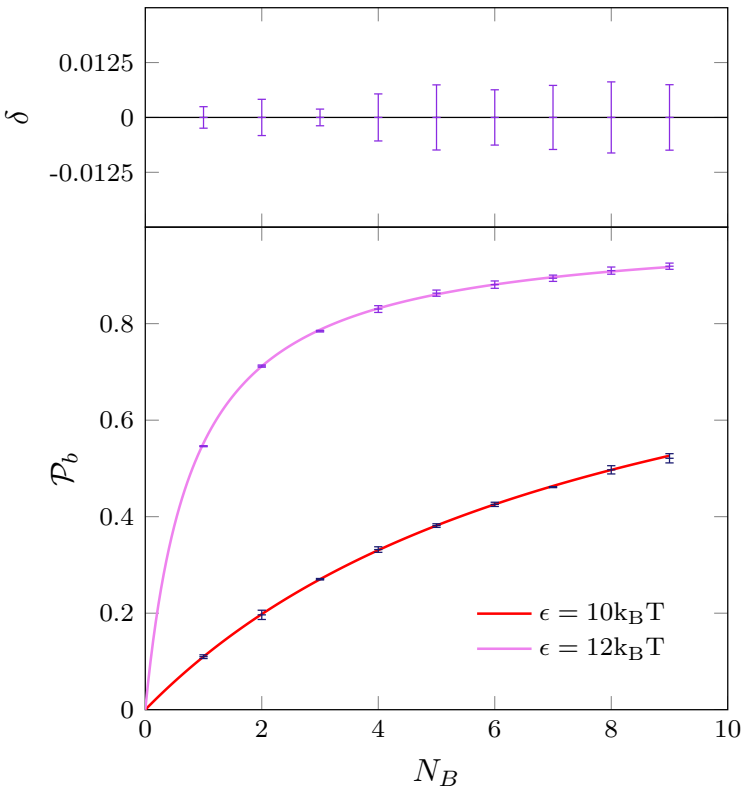


Figure 2.7: The probability $\mathscr{P}_b$ that a particle A is bound to a particle B, with increasing N_B in a box of constant volume V. In this case there is one A particle and N_B B particles in the box. The main panel shows $\mathscr{P}_b$ for two different values of the interaction strength ϵ. The symbols are the results of MD-GFRD and the solid lines are the mean-field prediction of Eq. 2.21. It is seen that the agreement is very good. The top panel shows the difference between the results of MD-GFRD and those of brute-force BD, $|\delta = \mathscr{P}_{b(MSS)} - \mathscr{P}_{b(BD)}|$, for $\epsilon = 12 k_B T$. Box width is 120σ, where σ sets the length scale of the Lennard Jones potential. The diffusion constant, which is not important for the value of $\mathscr{P}_b$, is $1 \mu m^2/s$.

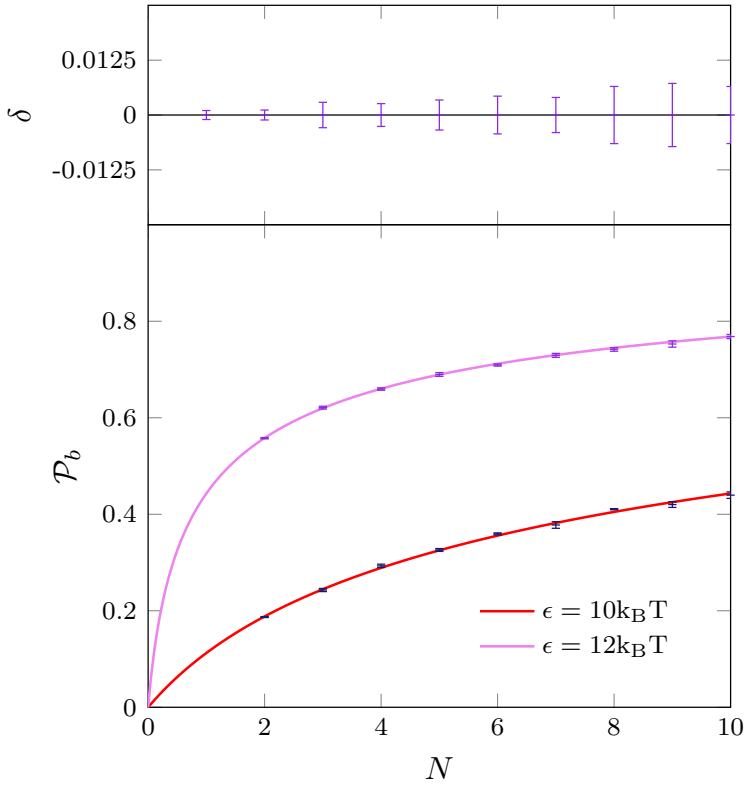


Figure 2.8: The probability $\mathscr{P}_b$ that a particle A is bound to a particle B, with increasing N_A and N_B in a box of constant volume V. In this case there are $N = N_A + N_B$ particles, with $N_A = N_B = \frac{N}{2}$. The main panel shows the results of MD-GFRD (symbols with error bars) together with the analytical prediction of Eq. 2.21 (sold lines). It is seen that the agreement between simulation and theory is very good. The top panel shows the difference between the MD-GFRD simulation results and those of brute-force BD, $|\delta = \mathscr{P}_{b(MSS)} - \mathscr{P}_{b(BD)}|$, for $\epsilon = 12 k_B T$. Clearly, the MD-GFRD results are in excellent agreement with those of brute-force BD. Box width is 120σ, where σ sets the length scale of the Lennard Jones potential. The diffusion constant, which is not important for the value of $\mathscr{P}_b$, is $1 \mu m^2/s$

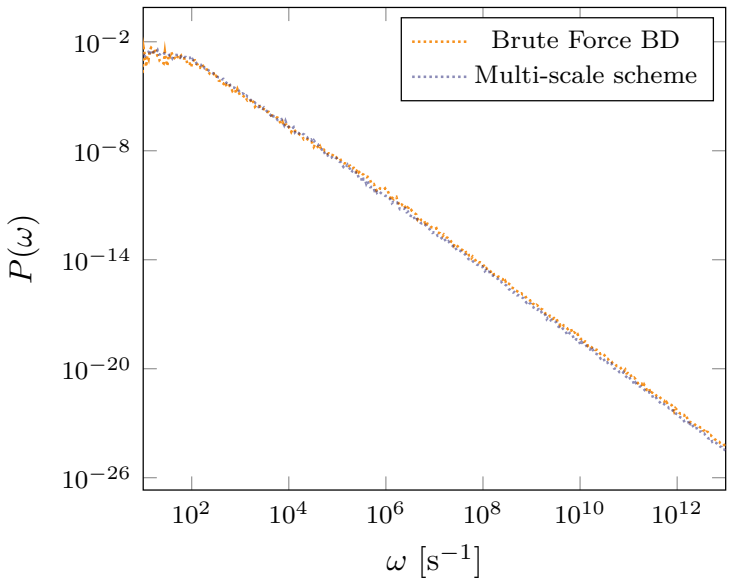


Figure 2.9: Power spectrum $P(\omega)$ of MD-GFRD validated with that of brute force BD proves that the scheme can simulate dynamic quantities. The simulation to generate the power spectrum was performed with one A and one B particle.

2.3.3. POWER SPECTRUM

MD-GFRD not only predicts the mean quantities, but also successfully describes dynamic quantities. To illustrate this we generate a power spectrum $P(\omega)$ of the time trace $n(t)$ of two particles switching from the bound state with $n(t) = 1$, to the unbound one with $n(t) = 0$. The Fourier transform of the correlation function of $n(t)$ gives the power spectrum. In Fig. 3.10 the power spectrum of MD-GFRD is compared to that of the brute force BD scheme. It can be seen that the two power spectra are in very good agreement with each other. This demonstrates the ability of MD-GFRD to predict dynamic quantities.

2.3.4. PERFORMANCE

One of the main motivations of developing MD-GFRD was the speed up GFRD provides when the particles are far apart. Unlike brute force Brownian dynamics which spends a lot of CPU time in propagating particles toward each other, GFRD makes large jumps in space and time when the particles are far apart on average. This, however, means that the efficiency of GFRD depends on the concentrations of the particles in the system. If the concentrations are low, the particles are far apart. This means that the GFRD domains will be large in size and the particles will be able to make large jumps in space and time. Here we can expect outstanding performance by the scheme, when compared to brute force BD. On the other hand, if the concentration is high, the particles are close to each

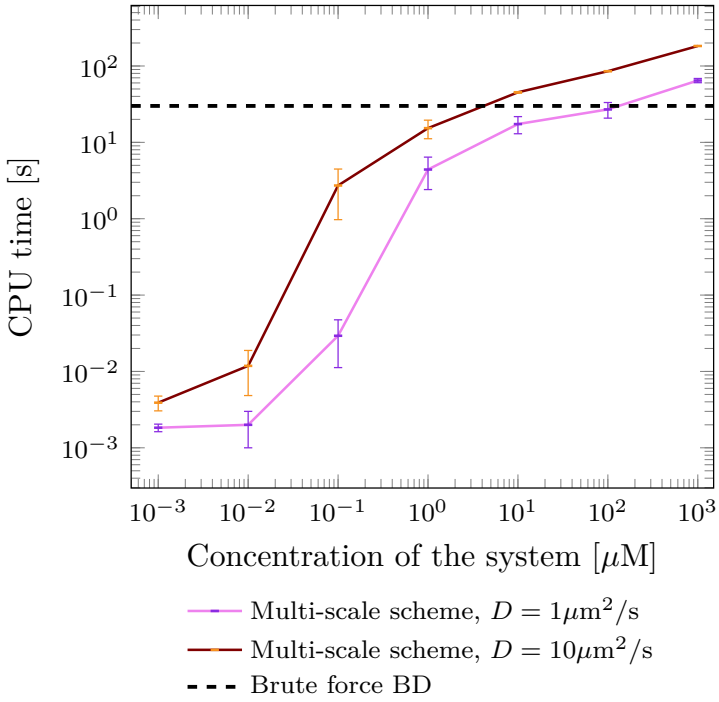


Figure 2.10: The CPU time to simulate 1ms real time as a function of the concentration of A and B, for MD-GFRD (solid line) and BD (dashed line). The concentration is varied by changing the volume of the simulation box, while the number of particles is kept constant at $N_A = N_B = 5$. It is seen that in the biologically relevant concentration range of nanomolar to micromolar the performance of MD-GFRD is much better than that of brute-force BD, but at higher concentrations the relative performance of MD-GFRD goes down. This is because at higher concentrations, the particles will be close to each other, and the system cannot capitalize on the potential of MD-GFRD to make large jumps in time and space.

other, the GFRD domains will be small, and the particles will not be able to make big jumps in space and time. Now we can expect some speed up, but not as high as the previous case. For even higher concentrations, we expect MD-GFRD to perform less than brute force BD because of the overhead of checking whether the scheme should switch between BD and GFRD.

In Fig. 3.11 we compare the performance of MD-GFRD to that of brute force BD. It should be noted that the BD algorithm used here is not optimized i.e. with neighbor lists etc. But since we use only up to ten particles, the speedup obtained by neighbor lists is not substantial. Fig. 3.11 shows that at low concentrations MD-GFRD can be much faster than brute-force BD, because in this regime the algorithm can exploit the power of GFRD to make large jumps in time and space. It is also seen that at higher concentrations, the relative performance decreases, because the particles are now closer to each other. In addition, the figure shows that the efficiency depends on the diffusion constant: the

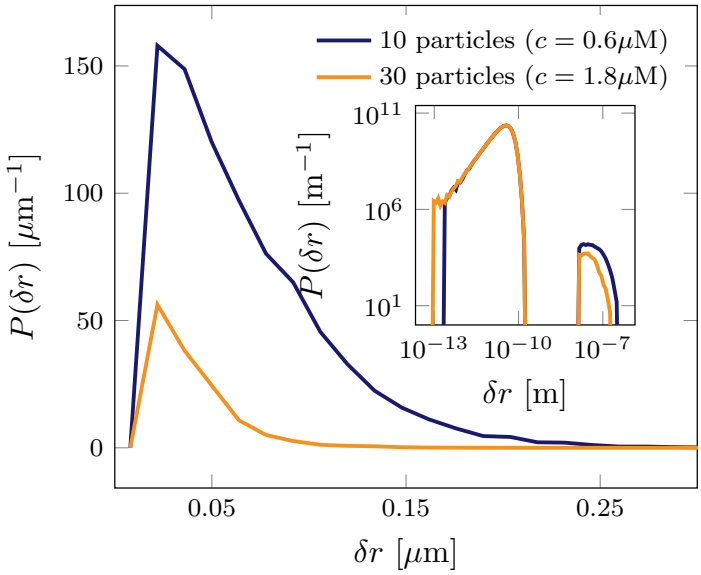


Figure 2.11: The probability density $P(\delta r)$ of spacial step sizes r for the multi-scale simulation, for two different number of particles N, yielding, at constant volume, two different concentrations c, as indicated. The inset shows the two regimes, GFRD and BD. It is seen that particles in the BD regime take a large number of smaller steps as compared to particles in the GFRD regime which take a smaller number of larger steps. Please also note that as the density is increased (by increasing the number of particles at constant volume), particles are more often propagated with BD. The simulations were performed with a box width of 120σ, well depth $\epsilon = 10k_B T$, diffusion constant $D = 1\mu\text{m}^2/\text{s}$.

lower the diffusion constant, the higher the performance of MD-GFRD as compared to brute-force BD. This can be understood by noting that at lower diffusion constants, the particles in the GFRD domains escape from their domains more slowly, and hence need to be updated less frequently.

From Fig. 3.11 we see that for a typical protein diffusion constant of $1\mu\text{m}^2/\text{s}$, MD-GFRD is less efficient than brute force BD for concentrations larger than around 1mM. This is in line with a previous performance test of GFRD (www.GFRD.org). Importantly, for cellular systems, mM concentrations are high: in gene networks the concentrations of transcription factors are typically in the nM regime, and in signal transduction pathways the concentrations tend to be in the μM range. As the figure shows, under these concentrations MD-GFRD is much faster than brute force BD.

2.3.5. SPACIAL STEP-SIZE

Fig. 2.11 shows the probability density of spacial step sizes in the multi-scale simulation for $D = 1\mu\text{m}^2\text{s}^{-1}$, $N = 10$ and $N = 30$ particles in a cubic box of width $L = 120\sigma$ (as before), corresponding to concentrations of 0.6μM and 1.8μM respectively. The inset

shows the same data on a log-log scale. This inset reveals two distinct regimes: one corresponding to BD particles with a step size of approximately 10pm and another regime corresponding to particles propagated with GFRD. This distribution has a peak at around 10nm, but exhibits a long tail, which means that steps of a 100nm are not uncommon, when the concentrations are low. This is the origin of the power of MD-GFRD.

2.4. CONCLUSIONS AND FUTURE WORK

We have developed a multi-scale scheme, called MD-GFRD, which can simulate reaction-diffusion systems at different length and time scales. When the particles are far apart, the scheme capitalizes on the power of GFRD. GFRD breaks down the N-particle problem into sets of one and two particle problems by putting single particles and pairs of particles into protective domains, and solve the reaction-diffusion problem for each of the domains analytically. MD-GFRD presented here only uses the Single domains of GFRD, which consist of single particles only. While the Pair domains provide speed up in the original GFRD scheme, much, if not most of the speed up, comes from the Single domains. For these Single domains, the Green Functions can be computed efficiently, we do not need to construct separate domains for the center-of-mass and for the inter-particle vector as for the Pair domains, and, most importantly, also the Single domains allow particles to make large jumps in time when they are far apart from each other. Hence, also in MD-GFRD, the time needed to propagate the particles towards one another is drastically reduced.

Since we want to describe the particle encounters with microscopic detail, MD-GFRD switches to a microscopic simulation technique when the particles come close together. For purpose of illustration, the MD-GFRD presented here employs Brownian Dynamics. However, to simulate the particle encounters, in principle any microscopic technique could be used, such as Molecular Dynamics or a coarse grained representation of MD, like Markov state modeling. All these techniques allow for a more accurate description of the orientational and conformational dynamics of the particles at the microscopic scale, which can influence the dynamics at the macroscopic scale [8].

MD-GFRD not only predicts average quantities, but also dynamic quantities successfully. Moreover, while it captures the particle dynamics at the microscopic level, it is much more efficient than brute force BD at low concentrations, as observed in cellular systems. MD-GFRD is a self-organizing algorithm that automatically zooms in on the particles that are close together, thus keeping particles as much as possible at the mesoscopic level, where they can be simulated with high speed.

Finally, MD-GFRD is very generic: it can be used to simulate any reaction-diffusion system at the particle level. It can not only simulate on a molecular level a wide range of cellular processes, but can also be used to describe the kinetics of self-assembly processes in polymeric systems, catalysis in a complex environment(eg. fuel cells), the dynamics of micellar surfactant systems or crystallization from dilute solution. In future work, we will include orientational dynamics into BD, where the orientation of a particle influences the behavior of the system.

ACKNOWLEDGEMENTS

We thank Tom Ouldridge for stimulating discussions and help with implementing the LD integrator. 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.

BIBLIOGRAPHY

[1] Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst Biol (Stevenage)* 1:230–236.

[2] Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.* 1:137–151.

[3] Lipková J, Zygalakis KC, Chapman SJ, Erban R (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. *J. Appl. Math.* 71:714–730.

[4] Flegg MB, Chapman SJ, Erban R (2012) The two-regime method for optimizing stochastic reaction-diffusion simulations. *J. R. Soc. Interface* 9:859–868.

[5] Schöneberg J, Noé F (2013) ReaDDy - A Software for Particle-Based Reaction-Diffusion Dynamics in Crowded Cellular Environments. *PLoS ONE* 8:e74261.

[6] Johnson ME, Hummer G (2014) Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. *Phys. Rev. X* 4:031037.

[7] Gillespie DT, Seitaridou E, Gillespie CA (2014) The small-voxel tracking algorithm for simulating chemical reactions among diffusing molecules. *J. Chem. Phys.* 141:234115.

[8] Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *P. Natl. Acad. Sci. USA* 107:2473–2478.

[9] Mugler A, Bailey AG, Takahashi K, Rein Ten Wolde P (2012) Membrane clustering and the role of rebinding in biochemical signaling. *Biophys. J.* 102:1069–1078.

[10] Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Bio.* 1:137–151.

[11] Van Zon JS, Rein Ten Wolde P (2005) Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.* 94:1–4.

[12] Prinz Jh, et al. (2011) Markov models of molecular kinetics : Generation and validation. *J. Chem. Phys.* 174105.

[13] Chodera JD, Singhal N, Pande VS, Dill Ka, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:1–17.

[14] Davidchack R, Ouldridge T, Tretyakov M (2015) New langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* 142.

[15] Nielsen SO, Moore PB, Ensing B (2010) Adaptive multiscale molecular dynamics of macromolecular fluids. *Phys. Rev. Lett.* 105:237802.

[16] Praprotnik M, Delle Site L, Kremer K (2005) Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* 123.

[17] van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.* 123:234910.

[18] Opplestrup T, Bulatov V, Gilmer G, Kalos M, Sadigh B (2006) First-Passage Monte Carlo Algorithm: Diffusion without All the Hops. *Phys. Rev. Lett.* 97:230602.

[19] van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762.

[20] Hummer G (2004) From transition paths to transition states and rate coefficients. *J. Chem. Phys.* 120:516–523.

[21] Allen RJ, Warren PB, ten Wolde PR (2005) Sampling Rare Switching Events in Biochemical Networks. *Phys. Rev. Lett.* 94:018104.

[22] Schuette C, Noe F, Lu J, Sarich M, Vanden-Eijnden E (2011) Markov state models based on milestoning. *J. Chem. Phys.* 134:204105.

3

MULTISCALE SIMULATIONS OF ANISOTROPIC PARTICLES COMBINING MOLECULAR DYNAMICS AND GREEN'S FUNCTION REACTION DYNAMICS

The modeling of complex reaction-diffusion processes in, for instance, cellular biochemical networks or self-assembling soft matter can be tremendously sped up by employing a multiscale algorithm which combines the mesoscopic Green's Function Reaction Dynamics (GFRD) method with explicit stochastic Brownian, Langevin, or deterministic Molecular Dynamics to treat reactants at the microscopic scale [A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde, J. Chem. Phys. **143**, *21: 214102 (2015)]. Here we extend this multiscale MD-GFRD approach to include the orientational dynamics that is crucial to describe the anisotropic interactions often prevalent in biomolecular systems. We present the novel algorithm focusing on Brownian Dynamics only, although the methodology is generic. We illustrate the novel algorithm using a simple patchy particle model. After validation of the algorithm we discuss its performance. The rotational Brownian Dynamics MD-GFRD multiscale method will open up the possibility for large scale simulations of e.g. protein signalling networks.*

3.1. INTRODUCTION

Complex systems such as biochemical networks in living cells, catalytic reactions in, e.g. a fuel cell, surfactant/water/oil mixtures, or self-assembling soft matter, can be modeled efficiently as reaction-diffusion systems. In such reaction-diffusion systems the spatial distribution of reactants and the stochastic nature of their interactions are crucial for the system's macroscopic behaviour. At sufficiently low concentrations, the time taken for the reactants to diffuse and randomly find each other is much larger than the time required for the reaction. For example, in cellular systems, the concentrations of proteins are often in the nM – μM range. Experiments indicate that proteins inside the living cell move by normal diffusion [1] with effective diffusion constants in the $1-10\mu m^2 s^{-1}$ range. This means that, with typical protein cross sections of 10nm, the time it takes for reactants to find each other is on the order of milliseconds to seconds. This is often much longer than the microsecond timescales on which the actual association events occur once the particles have found each other [2, 3]. Reaction-diffusion systems thus often exhibit a strong separation of length and time scales, with the diffusive search process happening on length and timescales of microns and milliseconds to seconds, and the reactions occurring on scales of nanometers and sub-milliseconds [3]. Simulating such systems with conventional, brute-force simulation techniques is notoriously difficult. Indeed, simulating cellular biochemical networks with straightforward brute-force Brownian Dynamics (BD) [4, 5, 6, 7] often means that most CPU time is spent on propagating the particles towards one another [8]. To overcome the inefficiency of straightforward BD requires special techniques such as Green's Function Reaction Dynamics (GFRD) [9, 10].

GFRD is a mesoscopic technique that decomposes the many particle reaction diffusion problem into sets of one- and two-body problems that can be solved analytically. This is achieved by putting single particles and pairs of particles in so-called *protective* domains that do not overlap with each other. For each of these domains the reaction-diffusion problem is solved analytically using Green's functions. This yields for each domain a next event *type* which can either be a reaction in the domain or an escape from the domain, as well as a next event *time*, i.e. the time at which this event occurs. These events are put in a scheduler list which is updated chronologically. This makes GFRD an asynchronous, event-driven algorithm. Since stochastic processes in the individual domains are independent of each other, GFRD is an exact algorithm to simulate large reaction-diffusion systems. As the particles make huge leaps in space and time in GFRD the computational effort in propagating the particles to one another is greatly reduced, making GFRD orders of magnitude faster than brute force BD. However, the particles are assumed to be idealized spheres interacting via an isotropic potential and the reactions to occur according to intrinsic rates in pair domains. Solving the Green's function for reactive events involving the complex anisotropic potentials required for proper modeling of proteins or other molecules is extremely cumbersome, and in fact most likely will reduce the efficiency of the GFRD approach substantially. In contrast, straightforward BD is able to naturally simulate orientational dynamics of protein particles with complex anisotropic (effective) interactions.

This observation raises the question whether it is possible to combine the computational power of GFRD with the microscopic detail of BD. In previous work, we introduced

a novel multi-scale scheme, called Molecular Dynamics-GFRD (MD-GFRD), which combines GFRD with a microscopic simulation technique such as deterministic molecular dynamics (MD), or stochastic Langevin Dynamics (LD) or Brownian Dynamics (BD)[11]. In this scheme GFRD handles diffusion of particles at the mesoscopic scale, while MD, LD or BD treats the particles that are coming close to each other. In previous work and here, we limit ourselves to BD, although the scheme can very easily formulated for MD and LD. The multi-scale algorithm defines the micro- and mesoscopic regions adaptively on the fly and switches seamlessly between the two techniques based on predefined scenarios.

In this work we extend MD-GFRD to incorporate the orientational dynamics of particles that interact via an anisotropic potential. As in the original MD-GFRD technique [11], GFRD is used for propagating the particles towards one another when they are far apart. Once the particles are within a predefined threshold distance from each other, the algorithm switches to BD. The complex orientational dynamics once the particles are close together is thus simulated with BD. When the particles are bound, MD-GFRD could in principle continue to simulate these particles with BD. However, in many cases, and typically in cellular systems, the particles are bound much longer than the time it takes to diffuse and thermalise within the interaction well, meaning that dissociation is a rare event. MD-GFRD exploits this separation of timescales by treating the dissociation as a first order reaction, with an intrinsic dissociation rate constant that has been pre-determined. After dissociation, the particles can be propagated again with GFRD. Importantly, however, after dissociation the particles do not immediately loose their orientational memory, which means that they must be propagated with Green's Functions that do not only describe the translational dynamics of the particles, but also their orientational dynamics. In this paper, we describe in detail how the MD-GFRD scheme switches between MD and GFRD and how this switching depends on the translational and orientational dynamics of the particles. We also present the Green's Functions that allow GFRD to simulate the particles' orientational dynamics.

The remainder of the paper is organized as follows. In the methods section we first give an overview of the MD-GFRD algorithm. Then we describe how the algorithm simulates the diffusion of particles with rotational degrees of freedom, both for particles in BD and GFRD mode. We discuss how MD-GFRD handles the association-dissociation reactions, and we describe how it switches between BD and GFRD propagation. In many systems, including that studied here, dissociation is a rare event. This means that computing the intrinsic dissociation rate constant, as used by MD-GFRD, requires rare event methodology, like Transition Interface Sampling [12] and Forward Flux Sampling (FFS) [13]. Here, we briefly describe how we use FFS to pre-compute the dissociation rate constant. We then illustrate the new technique by simulating the association and dissociation of patchy particles. In many cases, globular proteins can be coarse-grained as so-called *patchy particles*, where the complex binding sites are modeled as patches on a spherical particle. These patchy particles also play an important role in the modeling of soft matter[14, 15]. We demonstrate that the algorithm reproduces quantities that can be obtained analytically such as the equilibrium constants, binding probabilities and the power spectra of the binding reactions. We end with a discussion of the performance of the algorithm.

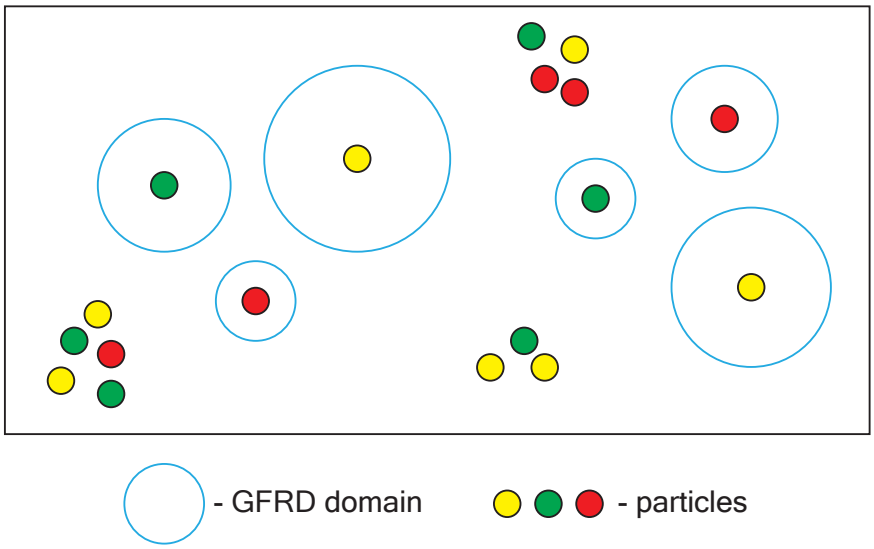


Figure 3.1: MD-GFRD scheme: particles that are far away other particles are put into a GFRD domain. For each GFRD domain the next event *time* and *type* is determined. These next event times are added to a chronologically ordered event list, and updated when the simulation time has reached the time of the next event. Particles that are close to other particles are propagated collectively with Brownian Dynamics.

3.2. METHODS

3.2.1. SUMMARY OF MULTISCALE APPROACH

The MD-GFRD algorithm is a generic algorithm that enables simulation of any reaction-diffusion system at the particle level. It allows for mono-molecular reactions of the type $A \rightarrow B + C + \ldots$ and bi-molecular reactions of the type $A + B \rightarrow C + D + \ldots$. By combining these two reactions, any complex biochemical network can be simulated. Here, however, we will limit ourselves to simple association-dissociation reactions $A + B \leftrightarrows C$. The MD-GFRD algorithm distinguishes two types of particles as shown in Fig. 3.1: 1) BD particles that are propagated collectively in a conventional, brute-force manner using small time steps, and 2) GFRD particles that are updated asynchronously in an event-driven manner Single particles that are sufficiently far away from all other particles according to a predefined cut-off distance are put into protective domains. For each of these domains, the algorithm determines, as in the conventional GFRD scheme [10], the next-event *type*, which is either a mono-molecular decay reaction (such as dissociation) or an exit of the particle from the domain, and the corresponding next-event *time*, which is when this next event will happen. The next-event times of the respective GFRD domains are put in a chronologically ordered event list, which is updated only when the simulation time has reached the time of the first next event. The event-driven nature of GFRD allows MD-GFRD to make large jumps in space and time when the domains are large. It is the

origin of the high efficiency of the scheme.

The other particles are simulated explicitly with BD. This part of the algorithm takes into account the forces between the particles when they come within the interaction range of the potential from each other. The BD propagation also naturally simulates the association reaction A + B → C: two particles A and B form the bound complex C when they enter the well of the interaction potential. The two monomers A and B in the dimer C could be propagated separately with BD, but it is more efficient to propagate them as a single particle C. The dissociation of C into A and B is then treated as a uni-molecular reaction event, which is added to the event list.

BD propagation is continued until one of the following events occurs: *i*) the escape of a particle from a GFRD domain; *ii*) the decay of a GFRD particle, e.g. the dissociation of C into A and B; *iii*) a BD particle dissociates into its products, e.g. the dissociation of C into A and B; *iv*) two BD particles A and B bind each other to form a dimer species C; *v*) a BD particle comes too close to a GFRD domain so that the GFRD domain must be burst, which means that a position for the particle in that domain is generated at the current simulation time; *vi*) BD particle moves sufficiently far away from all other BD particles and GFRD domains, so that it can be put into a GFRD domain. These six possible events are illustrated in Fig. 3.3. After the event has been executed, the system is updated accordingly; for newly formed GFRD domains, the next-event types and times are determined and inserted into the event list. The propagation of the BD particles is then resumed. The scheme becomes particularly powerful when most particles are in GFRD domains. A key objective is thus to keep the number of BD particles to a minimum.

The multiscale method that we pursue here involves particles interacting via anisotropic potentials. This requires an explicit BD integrator allowing rotational dynamics. Moreover, the GFRD part requires rotational Green's functions. In the next subsections we provide these ingredients, which constitute the most salient differences of the novel scheme with the previous isotropic MD-GFRD scheme [11]. In the subsequent subsection, we discuss in detail how the algorithm switches between GFRD and BD. The next two subsections describe how MD-GFRD handles the dissociation events and how the dissociation rate constant, needed in MD-GFRD, can be computed efficiently. In the last subsection, we describe the specific interaction potential used to illustrate how orientations can be included in MD-GFRD.

3.2.2. BROWNIAN DYNAMICS OF PATCHY PARTICLES

Brownian dynamics is used to simulate the solute particles at the microscopic scales. In this algorithm the position and the orientation of each solute particle in the BD regime is updated based on the total force and torque acting on the particle. The force and torque contain a deterministic component, which arises from the (solvent-mediated) interaction potential with the other solute particles and the frictional drag from the solvent, and a stochastic component, originating from the stochastic forces exerted by the solvent molecules. Although the interactions between particles are anisotropic, we model the particles as spheres of finite radius. We represent the rigid body orientation of the particles using a four component unit vector known as a quaternion, $q = (q_0, q_1, q_2, q_3)$.

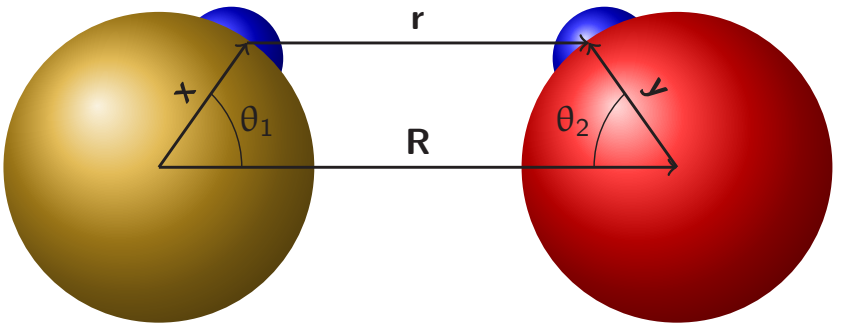


Figure 3.2: Each particle may have one or more attractive regions on its surface, called 'patches', that facilitate short ranged, highly directional attractive interactions.

The quaternion is an efficient encoding of the rotation matrix, A given by,

$$A = \begin{bmatrix} {q_0}^2 + {q_1}^2 - {q_2}^2 - {q_3}^2 & 2(q_1 q_2 + q_0 q_3) & 2(q_1 q_3 - q_0 q_2) \\ 2(q_1 q_2 - q_0 q_3) & {q_0}^2 - {q_1}^2 + {q_2}^2 - {q_3}^2 & 2(q_2 q_3 + q_0 q_1) \\ 2(q_1 q_3 + q_0 q_2) & 2(q_2 q_3 - q_0 q_1) & {q_0}^2 - {q_1}^2 - {q_2}^2 + {q_3}^2 \end{bmatrix}.$$

which relates vectors in the stationary lab frame, $\hat{u}_\mathrm{s}$, to the vectors in the moving body frame, $\hat{u}_\mathrm{b}$ via

$$\hat{u}_\mathrm{s} = A^T \hat{u}_\mathrm{b} \tag{3.1}$$

For example, the vectors $\hat{u}_\mathrm{b}$ might point to the patches on the surface of the particle which are fixed in the body frame.

Each particle has a center of mass, and one or more sticky spots on its surface called 'patches' (see Fig. 3.2). The particles interact with each other both via a center of mass isotropic pair potential and via a short ranged isotropic patch-patch interaction. We describe the anisotropic model potential that we employ for illustrative purposes in detail in Sec. 6.2.3. We note that the choice of potential is not limited to simple models. In principle any other anisotropic complex potential can be used, even an anisotropic protein-protein interaction derived from all atom MD simulations.

The particles are propagated with the first order Brownian dynamics integrator[16] explained in Algorithm 1:

Algorithm 2 The Brownian dynamics integrator [16] used in our multi-scale scheme. We consider n particles in three dimensions with center of mass coordinates $\mathbf{r} = (r^{1\top}, \ldots, r^{n\top})^\top \in \mathbb{R}^{3n}$, $r^j = (r_1^j, r_2^j, r_3^j)^\top \in \mathbb{R}^3$, and rotational coordinates in the quaternion representation $\mathbf{q} = (q^{1\top}, \ldots, q^{n\top})^\top$, $q^j = (q_0^j, q_1^j, q_2^j, q_3^j)^\top \in \mathbb{S}^3$, such that $|q^j| = 1$. Particles are characterized by their mass m, the mass moment of inertia $M = \frac{8}{15} m\sigma^2$, and the translational and rotational friction coefficients, γ and Γ, respectively. These parameters can differ among species. Note that this algorithm uses γm as an effective friction. The mass in this effective friction is canceled by the presence of the mass in the integration itself. Furthermore, δt is the time-step used in

the simulations, $\beta = \frac{1}{k_B T}$, ξ_k and $\eta_k^{j,l}$ are independent and identically distributed (i.i.d.) Gaussian random variables. **f** is the total force and $\mathbb{F}$ is the total torque, which follow from the interaction potential.

$$\mathbf{R_0} = \mathbf{r},\ \mathbf{Q_0} = \mathbf{q},\ |q^j| = 1,\ j = 1,\ldots,n,$$

$$\mathbf{R}_{k+1} = \mathbf{R}_k + \frac{\delta t}{\gamma m}\mathbf{f}(\mathbf{R}_k, \mathbf{Q}_k) + \sqrt{\delta t}\sqrt{\frac{2}{\gamma\beta m}}\xi_k,$$

$$Y_k^j = \frac{\delta t}{\Gamma M}\mathbb{F}_j(\mathbf{R}_k, \mathbf{Q}_k) + \sqrt{\delta t}\sqrt{\frac{2}{\Gamma\beta M}}\Sigma_{l=1}^{3}\eta_k^{j,l}S_l,$$

$$Q_{k+1}^j = \exp(Y_k^j)Q_k^j.$$

where S_1, S_2 and S_3 are 4×4 matrices $S_1 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$, $S_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $S_3 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

3.2.3. GREEN'S FUNCTIONS FOR ROTATIONS

GFRD handles the free diffusion of single particles. A freely-moving particle will undergo rotational as well as translational diffusion. Although the interactions between particles are anisotropic, we model the particles as spheres of finite radius for the purpose of modeling diffusion. This assumption allows the decoupling of the rotational and translational diffusion of isolated particles, which is possible since in MD-GFRD the GFRD domains only contain single particles. The Green's functions for translational diffusion are given by the Green's functions for single particles inside Single GFRD domains, detailed in previous work [11]. These Green's functions determine (probabilistically) when the particles escape from their respective domains, or what their radial positions inside the domains become when the domains are burst. Although rotational motion does not influence the center-of-mass dynamics of a freely diffusing particle, and hence cannot cause escape from Single Domains, it is nonetheless important to reproduce the decorrelation of orientations for particles evolving under GFRD. For example, simply drawing orientations at random when a particle leaves a GFRD Single Domain will lead to unphysically rapid decorrelation of orientations when domains are short-lived, and influence properties such as rebinding probability.

More specifically, on bursting or escape from a Single Domain, a new orientation Ω is drawn using the Green's function $G(\Omega, \Omega_0, t)$, with Ω_0 being the initial orientation and t the time since domain formation. The Green's functions, expressed in terms of Euler angles α, β, γ, can be found in the literature [17, 18, 19]. For particles with spherically symmetric diffusion tensors the Green's function is

$$G(\alpha,\beta,\gamma,\alpha_0,\beta_0,\gamma_0,t) = \sum_{L=0}^{\infty}\sum_{K,M=-L}^{L}\frac{2L+1}{8\pi^2}D_{K,M}^{(L)*}(\alpha_0,\beta_0,\gamma_0)D_{K,M}^{(L)}(\alpha,\beta,\gamma)\exp(-D_r L(L+1)t). \tag{3.2}$$

Here, D_r is the threefold degenerate eigenvalue of the diffusion tensor, given by $D_r = k_B T/(8\pi\eta R^3)$ for a particle of radius R in a fluid of viscosity η. The quantities $D_{K,M}^{(L)}(\alpha,\beta,\gamma)$ and its complex conjugate $D_{K,M}^{(L)*}(\alpha,\beta,\gamma)$ are elements of the Wigner rotation matrices

[17, 18, 19]:

$$D^{(L)}_{K,M}(\alpha,\beta,\gamma) = \exp(-iK\alpha)\, d^{(L)}_{K,M}(\beta) \exp(-iL\gamma), \tag{3.3}$$

with

$$d^{(L)}_{K,M}(\beta) = ((L+K)!(L-K)!(L+M)!(L-M)!)^{1/2} \times \sum_{S=\max(0,M-K)}^{\min(L+M,L-K)} \left(\frac{(-1)^{K-M+S}.[\cos(\beta/2)]^{2L+M-K-2S}[\sin(\beta/2)]^{K-M+2S}}{(L+M-S)!S!(K-M+S)!(L-K-S)!} \right) \tag{3.4}$$

For the purposes of clarity, we emphasize that the Euler angles used here should be understood in the following way. If a body frame B has an orientation $\Omega = (\alpha, \beta, \gamma)$ with respect to some reference frame F, then B can be obtained from F by:

1. Rotating F around F_z by γ to give F'.
2. Rotating F' around F_y by β to give F''.
3. Rotating F'' around F_z by α to give B.

Moreover, note that the Green's functions are defined without the Jacobian, so that (α, β, γ) should be drawn from the distribution $\sin(\beta)G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$.

Drawing directly from such a distribution is non-trivial. However, rejection sampling can be used if the maximum of $\sin(\beta)G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$ is known. Physically, the most likely orientation is always aligned with the initial direction, which suggests a rejection scheme in which a trial orientation (α, β, γ) is drawn uniformly from $([0, 2\pi], [0, \pi], [0, 2\pi])$, and accepted with a probability

$$\frac{\sin(\beta)G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)}{\sin(\beta_0)G(\alpha_0, \beta_0, \gamma_0, \alpha_0, \beta_0, \gamma_0, t)},$$

with Euler angles defined with respect to the lab frame. Unfortunately, the angular Jacobian implies that $\sin(\beta)G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$ is not in general maximized by $(\alpha = \alpha_0, \beta = \beta_0, \gamma = \gamma_0)$, violating a requirement of rejection sampling. It is true, however, that $\sin(\beta)G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$ is maximized by $(\alpha = \alpha_0, \beta = \beta_0, \gamma = \gamma_0)$ if $\beta_0 = \pi/2, \alpha_0 = 0, \gamma_0 = 0$. We therefore define a new reference frame F_{temp} for each calculation such that the particle initially has orientation $\Omega_0 = (0, \pi/2, 0)$ with respect to F_{temp}. Using rejection sampling, we can then obtain a new orientation $\Omega = (\alpha, \beta, \gamma)$ with respect to F_{temp}. The particle orientation is updated by first rotating the particle by $-\pi/2$ about the z-axis of the original particle frame to obtain a particle aligned with F_{temp}, and then performing rotations (α, β, γ) about the axes of F_{temp} as outlined above.

Even with rejection sampling, drawing from the distribution can be computationally challenging due to the costs of evaluating Green's functions. Eq. 3.2 has an infinite sum that must be truncated; we perform truncation when new contributions are smaller than the current value by a factor of 10^8. To reduce the cost of the summations, we find it helpful to tabulate factorials. We also note that terms in Eq. 3.2 can be combined in complex conjugate pairs to eliminate imaginary numbers during the calculation.

Accurate evaluation of the Green's function is most challenging when $D_r t < 1$, when $G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$ is sharply peaked and many terms are needed. For small $D_r t$, we use early rejection, discarding a large fraction of draws of (α, β, γ) if $(\alpha, \beta - \beta_0, \gamma)$ is large without evaluating $G(\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, t)$, and compensating for this bias at the acceptance stage. Finally, for values of $D_r t < 0.05$, we use the approximate approach of rotating about a random axis through an angle $\phi = \sqrt{\phi_x^2 + \phi_y^2 + \phi_z^2}$, where ϕ_i are i.i.d. random variables drawn from a Gaussian of mean 0 and variance $2D_r t$ [20].

3.2.4. HANDLING THE DISSOCIATION/ASSOCIATION REACTION IN MD-GFRD

While particles that are sufficiently far away from each other can be propagated with GFRD, particles that are within a pre-defined cutoff distance from each other will be propagated with MD, or, as we restrict ourselves to here, BD. As described in more detail in the next section, this cut-off distance is beyond the range of the interaction potential, r_c. Indeed, the association between two particles, which is driven by their intermolecular attraction forces, is thus simulated explicitly with BD. Also the dissociation reaction could in principle be simulated with BD: we could explicitly simulate the bound monomers in the dimer A-B, until they dissociate again into A and B. However, the bound state is typically very stable: the time the particles spent inside the potential well is typically much longer than the time it takes for the particles to loose their orientation and thermalise inside the well. Simulating these particles explicitly means that much CPU time would be wasted on propagating them while they simply rattle around each other inside the potential well. In MD-GFRD, we therefore exploit that dissociation is a rare event: when two BD particles meet a predefined criterion signifying that they are deep inside the interaction well, the two 'reactants' A and B are replaced by species C. In turn, the dissociation of C into A and B is treated as a first-order reaction $\mathrm{C} \rightarrow \mathrm{A} + \mathrm{B}$ with a dissociation rate constant k_d.

More specifically, when two BD particles come within a distance such that their interaction energy E drops below some predefined threshold E_bind, then the particles A and B are replaced by a single particle of species C, with a position that is given by the center-of-mass of the reactants A and B. If space permits, the C particle is directly put into a GFRD domain, which significantly speeds up the simulation. If there is no space to construct a protective domain, the C particle is propagated with BD. The C particle then diffuses, either explicitly with BD or implicitly with GFRD, until it dissociates again into the monomers A and B at a later time τ_d. Since the interaction well is deep, τ_d will be exponentially distributed:

$$q_\mathrm{d}(t) = k_\mathrm{d} e^{-k_\mathrm{d} t}. \tag{3.5}$$

Knowing the dissociation rate constant k_d, the time τ_d can thus be sampled from

$$\tau_\mathrm{d} = -k_\mathrm{d} \ln(\mathscr{R}_\mathrm{d}), \tag{3.6}$$

where $\mathscr{R}_\mathrm{d} \in [0, 1]$ is a uniformly distributed random number.

The intrinsic dissociation rate constant k_d could in principle be inferred from experiments. However, a more consistent and rigorous approach is to obtain k_d from a simulation that is performed prior to the MD-GFRD simulation of interest. This pre-simulation can then also be used to generate the distributions of the positions and orientations of

A and B at the moment of dissociation. In the MD-GFRD simulation, the positions and orientations of the particles at the moment of dissociation can then be sampled from these distributions, respectively.

In our previous study on isotropic potentials, we determined k_d by performing a brute force BD simulation of two particles prior to the MD-GFRD simulations [11]. However, the particles in our model interact via an anisotropic interaction potential. This anisotropic interaction is mediated via patches on the surfaces of the particles, see Fig. 3.2. The range of the patch-mediated interaction must be short, in order to provide a strong anisotropy in the interaction. The short range, however, means that the well of the patch-mediated potential must be deep in order to induce significant binding: the depth of the well, $\sim 25 k_B T$, is much larger than that of isotropic particles, $\sim 5 k_B T$. The deep well makes it very hard to obtain good statistics in determining the distribution of dissociation times via brute force simulations. However, it is possible to efficiently compute the dissociation rate with rare event techniques such as Transition Interface Sampling [12] or Forward flux sampling (FFS)[13]. Here we use the latter technique, which we describe in section 3.2.6.

3.2.5. COUPLING BD AND GFRD

Now that we have described how MD-GFRD simulates the association and dissociation of two particles A and B, we will discuss how the algorithm switches between BD and GFRD when simulating many particles. At any one point in time, the simulation consists of a set of isolated particles inside GFRD domains that each have a radius of at least d_{min}, and a set of particles that are propagated with BD and interact with each other via a pair potential that has an interaction range r_c. There is also a chronologically ordered next-event list that contains the times at which the GFRD particles escape from their respective domains, and the times at which the respective particles dissociate, be they in GFRD or BD mode. The particles that are not inside GFRD domains are propagated with brute-force BD until the first next-event happens. This event can be an event from the next-event list, but it can also be the formation of a GFRD domain or the bursting of a GFRD domain when a BD particle comes too close it. After the event has been executed, BD propagation is resumed.

Specifically, before each step of BD propagation, the algorithm checks for the following events, as illustrated in Fig. 3.3:

ESCAPE FROM A GFRD DOMAIN

When the next event in the list is a particle that escapes from a single domain, that particle is put at a random center of mass position on the surface of the domain, with an orientation sampled from Eq. 3.2. The domain is removed and the particle is put in BD mode. This event is shown in Fig. 3.3.I. Note that at the next BD time step, the algorithm will check whether the particle can be put into a protective GFRD domain again (see Sec. 3.2.5).

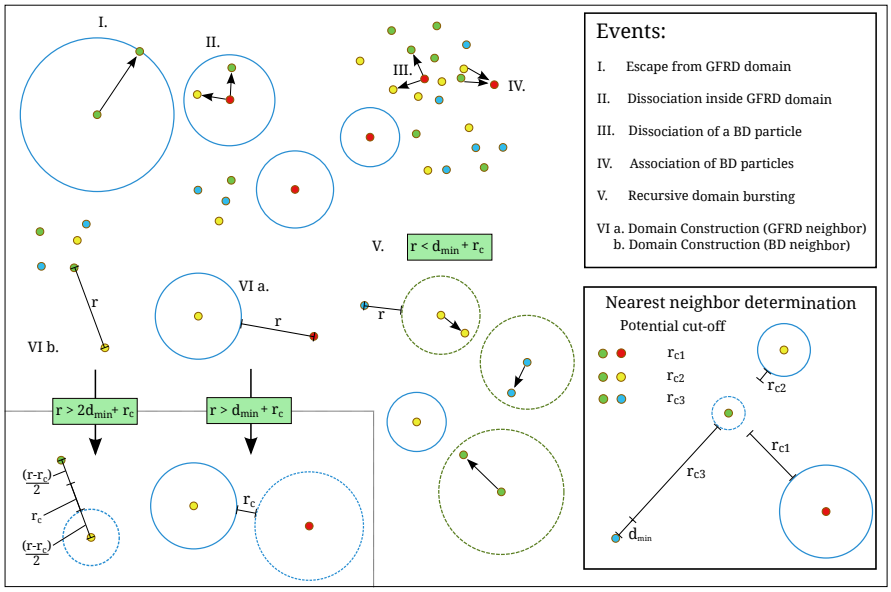


Figure 3.3: At each BD time step, the algorithm checks whether BD should be interrupted. The BD propagation is halted when the time of the first next event occurs before the global simulation time at the end of the time step. These next events can be any of the following: I. A particle escapes from a GFRD domain; the position of the particle is updated to a randomly chosen point on the surface of the domain and the domain is removed. II. A particle dissociates inside a GFRD domain; the domain bursts and the particle is updated to a position and orientation sampled using Green's functions, and is then replaced by its product particles. III. A BD particle dissociates; it is replaced by its product particles. IV. The binding energy of two BD particles is below the binding threshold; the particles enter the bound state and are replaced by a single product particle. V. The distance from a BD particle to a domain is smaller than $d_{\mathrm{min}} + r_{\mathrm{c}}$; the neighboring domain is burst and the position and orientation of the particle in this domain is updated. This particle may in turn burst another domain and this happens recursively until there is no BD particle within a distance $d_{\mathrm{min}} + r_{\mathrm{c}}$ from any other domain. VI. a. The distance between a BD particle and its nearest neighbor is larger than $d_{\mathrm{min}} + r_{\mathrm{c}}$ in case the nearest neighbor is a GFRD domain; a domain of radius $r - r_{\mathrm{c}}$ is built on the BD particle. b. The distance between a BD particle and its nearest neighbor is larger than $2d_{\mathrm{min}} + r_{\mathrm{c}}$ in case the nearest neighbor is a BD particle; a domain of radius $0.5(r - r_{\mathrm{c}})$ is built on the BD particle of interest. The inset shows the procedure for determining the nearest neighbor, which is the GFRD domain or the BD particle with the closest interaction horizon to the (central green) particle of interest: for BD particles the relevant distance is the distance minus the sum of the minimum domain radius d_{min} and the potential interaction range r_c, while for a GFRD domain the relevant distance is the distance to the surface of that domain minus r_c. In the example configuration the blue particle is the nearest neighbor.

DISSOCIATION INSIDE A GFRD DOMAIN

When the next event is a particle C inside a GFRD domain decaying into its products at time t, the domain is burst and a new radial position r for the reactant is generated according to the normalized translational Green's function $p(r,t,|r_0,t_0)/S(t-t_0|r_0)$, where r_0 is the original position of the particle, which is the center of the domain constructed at time t_0, and $S(t-t_0|r_0)$ is the survival probability. The reactant is replaced by its products, whose configuration is chosen at random from the ensemble of configurations recorded at the moment of dissociation, obtained in the FFS pre-simulation. This event is shown in Fig. 3.3.II.

DISSOCIATION OF A BD PARTICLE

When the next event is the dissociation of a BD particle, the particle is replaced by its products, whose configurations are chosen at random from the ensemble of configurations recorded at the moment of dissociation in the FFS pre-simulation. This event is shown in Fig. 3.3.III.

ASSOCIATION OF BD PARTICLES

When the pair potential energy between two particles becomes smaller than a threshold energy, here taken to be $E_{\mathrm{bind}} = -10k_{\mathrm{B}}T$, the two particles are defined to be in the bound state. The two particles are replaced by a single BD particle at their center of mass. This event is shown in Fig. 3.3.IV. Note that at the next BD step, the algorithm will check whether the particle can be put into a GFRD domain, as described under Sec. 3.2.5.

RECURSIVE DOMAIN BURSTING

When a BD particle comes at time t within a distance of $d_{\mathrm{min}} + r_{\mathrm{c}}$ from the surface of a GFRD domain, the domain is burst and a radial position r of the particle inside that domain is drawn from the normalized translational Green's function $p(r,t|r_0,t_0)/S(t-t_0|r_0)$, where t_0 is the time and r_0 the position of the center of the domain when it was constructed, and $S(t-t_0|r_0)$ the survival probability. A new orientation of the particles is sampled from Eq. 3.2. If this particle, after updating its position, comes within a distance of $d_{\mathrm{min}} + r_{\mathrm{c}}$ from another domain, that domain is also burst. This may lead to a cascade of domain bursting, which ceases when no BD particle is within a distance of $d_{\mathrm{min}} + r_{\mathrm{c}}$ from any GFRD domain. This event is shown in Fig. 3.3V. Note that domains are always at least r_c apart from each other.

DOMAIN CONSTRUCTION

For each BD particle, the algorithm determines the nearest neighbor, which is either another BD particle or a GFRD domain. The procedure to determine the nearest-neighbor distance depends on whether the neighbor is a BD particle or a GFRD domain, as shown in the inset of Fig. 3.3. A BD particle is put into a GFRD domain when the distance r between the particle and its nearest neighbor:

(a) is larger than $d_{\mathrm{min}} + r_{\mathrm{c}}$ in case the nearest neighbor is a GFRD domain. A domain of radius $(r - r_{\mathrm{c}})$ is built around the particle of interest. This event is shown in Fig. 3.3.VI a.

(b) is larger than $2d_{\min} + r_c$ in case the nearest neighbor is a BD particle. A domain of radius $0.5(r - r_c)$ is built around the particle of interest. This allows enough space to build a domain with a radius of at least $d_{\min}$ around the neighbor, thus preventing the neighbor from prematurely bursting the newly built domain. This event is shown in Fig. 3.3.VI b.

For the newly constructed domain the tentative next-event times for the respective tentative event types (e.g. dissociation and escape) are determined, and the event type with the smallest tentative next-event time is added to the event list. To achieve maximum efficiency, the minimal domain size $d_{\min}$ should be as small as is practical.

3.2.6. COMPUTING THE DISSOCIATION RATE WITH FORWARD FLUX SAMPLING

The Forward Flux Sampling (FFS) algorithm enables efficient evaluation of rare event kinetics. FFS uses a series of interfaces between the reactant and the product states to construct the transition path ensemble and calculate the corresponding transition rate. Each interface is defined by an order parameter λ: the reactant state is defined by $\lambda < \lambda_{-1}$ and the product state by $\lambda > \lambda_n$. The remaining interfaces are defined by intermediate values of λ: $(\lambda_0 \ldots \lambda_{n-1})$. The FFS technique requires that $\lambda_{i+1} > \lambda_i$ for all i, and all the trajectories from reactant to product state pass through each interface in succession as shown in Fig. 3.4. Trajectories starting in the reactant state and reaching product state are rare, but trajectories starting at an interface and crossing the next interface are more common. This is the central idea used in FFS [13].

Here we use the 'direct' FFS variant, DFFS, to compute the dissociation rate [21]. In this process the reactant state is the bound A,B dimer, and the product state corresponds to the dissociated dimer. For the purpose of simulating dissociation, we take an order parameter to determine the interfaces based on a combination of the energy of interaction and the inter-particle distance. The reactant bound state interface λ_{-1} is defined by a potential energy E_{bind}, while the product state is defined by zero potential energy in addition to an inter-particle distance larger than the cut-off r_c .

In the first step of FFS, a brute-force BD simulation is performed to compute the flux ϕ of crossing the interface λ_0 while coming from the bound state. This brute-force simulation generates an ensemble of configurations at λ_0. In the next step, a trajectory is fired from a randomly chosen configuration from this ensemble; this trajectory is then propagated until it either hits the next interface λ_1 or returns to the reactant state (i.e., recrosses λ_{-1}). This procedure is repeated until a sufficiently large number of configurations at the next interface λ_1 is generated. The fraction of trajectories that makes it from λ_0 to λ_1 yields the conditional probability $P(\lambda_1|\lambda_0)$ that a trajectory that comes from the bound state and crosses λ_0 for the first time will subsequently reach λ_1 instead of returning to the bound state. This whole procedure is then repeated for all subsequent interfaces until the final interface λ_n is reached, signifying the fully dissociated pair. Under the assumption of rare event kinetics, the intrinsic dissociation rate k_d is then given by [12, 13]

$$k_d = \phi \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i). \tag{3.7}$$

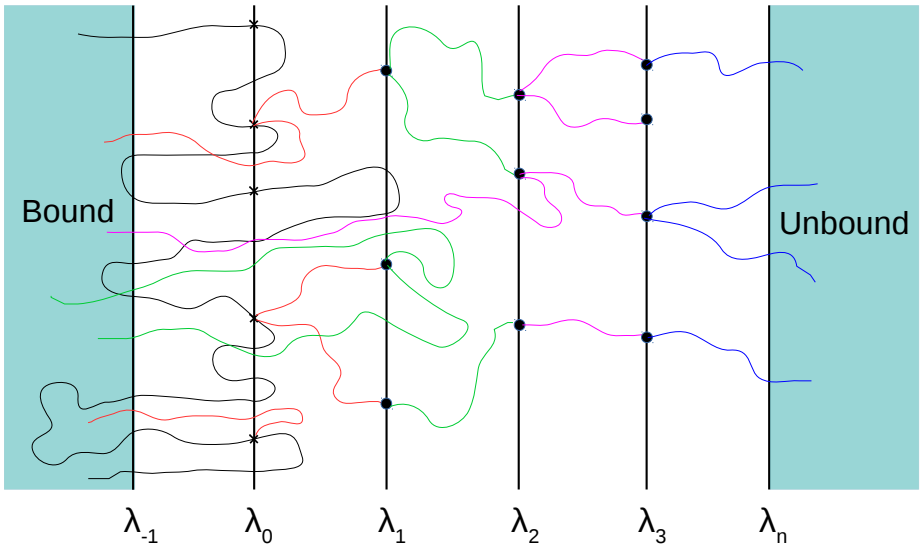


Figure 3.4: An illustration of the FFS method. An ensemble of transition paths is generated by starting trial runs from randomly picked configurations on interfaces, which are the end points of previous successful trial runs.

3.2.7. ILLUSTRATIVE ANISOTROPIC INTER-PARTICLE POTENTIAL

In this section we describe the interaction potential for the specific patchy-particle system. We reiterate that our multi-scale scheme is independent of the choice of potential, and can in principle be applied with arbitrarily complex potentials.

For convenience, we split our inter-particle potentials into three parts. Every pair of particles experiences a repulsive potential $U_{\text{rep}}(R)$ and an isotropic attractive potential $U_{\text{isoAtt}}(R)$ based on the distance R between the centers of mass. Additionally, each pair of complementary patches interacts through an attractive potential $U_{\text{att}}(r)$ based on the distance r between complementary patches (see Fig. 3.2). For a pair of particles with a single pair of complementary patches,

$$U(R,r) = U_{\text{rep}}(R) + U_{\text{isoAtt}}(R) + U_{\text{att}}(r). \tag{3.8}$$

Mediating the attractive interactions through surface-based patches naturally captures short-range contact interactions.

It is common to use 12-6 Lennard-Jones or related potentials in biomolecular modeling. Although the r^{-6} dependence is required for van der Waals interactions between atoms and even between larger entities, in general there is no fundamental reason to choose this functional form in case of complex effective interactions between biomolecules, e.g. hydrophobic interactions. In preliminary simulations, we observed that using Lennard-Jones potentials leads to numerical difficulties, forcing the use of extremely small time steps. The underlying reason is that Lennard-Jones potentials have a large curvature close to the minimum of the bound state, a situation for which the Brownian integra-

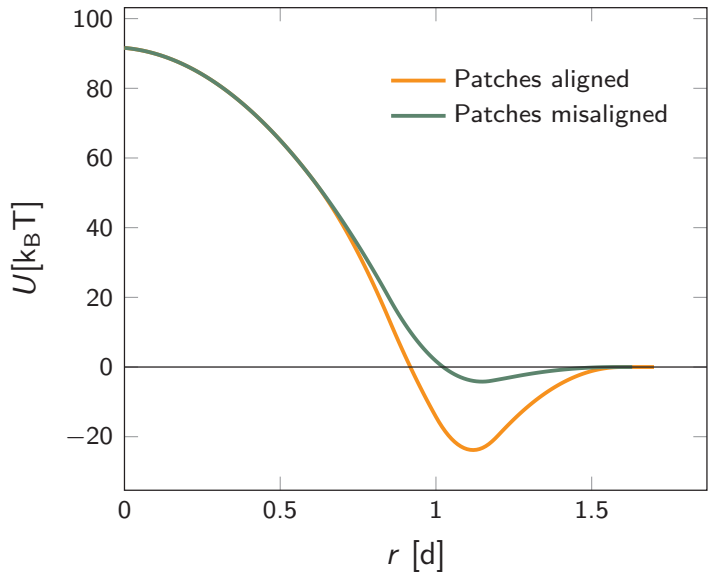


Figure 3.5: Inter-particle interactions. Total interaction potential $U_{\text{rep}}(R) + U_{\text{isoAtt}}(R) + U_{\text{att}}(R - 2d_{\text{patch}})$ for two particles with perfectly aligned complementary patches, and the total interaction potential $U_{\text{rep}}(R) + U_{\text{isoAtt}}(R) + U_{\text{att}}(R + 2d_{\text{patch}})$ when the complementary patches are completely misaligned. The existence of patches introduces an attractive bound state with the particles in close contact.

tor is poorly suited. This effect is exacerbated by the use of short-ranged anisotropic attractions between particles, which reduces the entropy of the bound state and must be compensated for by stronger attractive potentials, in order to model realistic equilibrium binding constants. Stronger attractive potentials lead to larger second derivatives of the potential. Moreover, requiring potentials to be short-ranged and orientation-specific implies variation over short length and angular scales, again increasing the second derivatives of the potential.

Instead of using a Lennard-Jones type potential we therefore illustrate our method using piece-wise quadratic potentials similar to those employed elsewhere [22]. These potentials give us more control over the shape, and allow for easier integration with potentials that are short-ranged and highly orientation-specific. We stress that using an alternative potential that is more challenging for the integrator would not remove the advantages of the multi-scale scheme.

$U_{\text{rep}}(R)$, $U_{\text{isoAtt}}(R)$ and $U_{\text{att}}(r)$ have the form

$$U_i(x) = \begin{cases} \epsilon_i(1 - a_i\left(\frac{x}{\sigma}\right)^2) & \text{if } x < x_i^\star, \\ \epsilon_i b_i(\frac{x_i^c}{\sigma} - \frac{x}{\sigma})^2 & \text{if } x_i^\star < x < x_i^c, \\ 0 & \text{otherwise,} \end{cases} \tag{3.9}$$

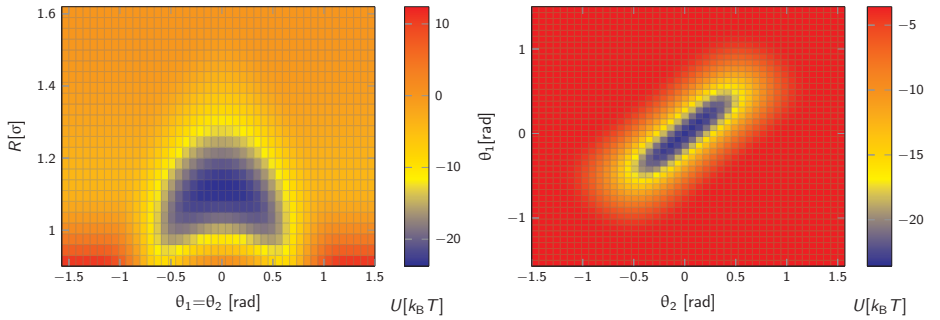


Figure 3.6: Total potential $U_{\rm rep}(R) + U_{\rm isoAtt}(R) + U_{\rm att}(r)$ for two particles with complementary patches. (a) Potential as a function of the distance between centres of mass R and alignment of patches with inter-particle vector, θ_1 and θ_2 (see definition in Fig. 3.2), given $\theta_1 = \theta_2$. (b) Potential as a function of θ_1, θ_2 given $R = 1.1\sigma$. Note the relatively narrow range of orientations over which strong bonding occurs.

with $i = \text{rep}, \text{isoAtt}, \text{att}$, respectively. The overall strength ϵ_i, the length scale σ (i.e. the particle diameter), 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 $x_i^\star$. For our illustrative purposes, we take the following parameters: $\epsilon_{\rm rep} = 100 k_{\rm BT}$, $a_{\rm rep} = 1$ and $R^*_{\rm rep} = 0.85\sigma$, implying $b_{\rm rep} = 2.6036$ and $R^c_{\rm rep} = 1.1764\sigma$; $\epsilon_{\rm att} = 20 k_{\rm B} T$, $a_{\rm att} = 20$ and $r^*_{\rm att} = 0.1\sigma$, implying $b_{\rm att} = 5$ and $r^c_{\rm att} = 0.5\sigma$; and $\epsilon_{\rm isoAtt} = 10 k_{\rm BT}$, $a_{\rm isoAtt} = 1$ and $R^*_{\rm isoAtt} = 0.85\sigma$, implying $b_{\rm rep} = 2.6036$ and $R^c_{\rm rep} = 1.1764\sigma$.

In Fig. 3.5, we plot the resulting total inter-particle potential as a function of distance R when the two complementary patches are aligned to face each other, so that $r = R - \sigma$. A narrow attractive well corresponding to the two particles being in close contact is evident. For comparison, we also show the total inter-particle potential as a function of R when the two complementary patches are misaligned to face opposite each other, so that $r = R + \sigma$. In this case, the patches do not contribute to the interaction; the non-specific, isotropic part of the potential, however, still gives rise to a weak attraction. In Fig. 3.6, we demonstrate the orientational dependence of the attractive potential, showing that the attractive interaction is highly sensitive to misalignment. We note that our choice of potential makes truncation at short distances relatively trivial. This is helpful in allowing rapid switching to GFRD domains once the particles are separated.

For our model potential the interaction range is set $r_{\rm c} = 1.6\sigma$, where the pair potential in Eq. 6.29 has vanished. Moreover, in the MD-GFRD simulations we set the minimum domain size $d_{\rm min} = 0.5\sigma$ and the particle diameter to $\sigma = 5$nm.

3.3. RESULTS AND DISCUSSION

We test the MD-GFRD simulation using the patchy-particle model described in section 6.2.3. In the simulations there are three species of particles, A, B and C, which react

according to

$$A + B \leftrightarrows C. \tag{3.10}$$

The system specific parameters of the simulation are as follows: The particle diameter is $\sigma = 5$nm, the time step $\delta t = 0.1$ns, the mass of the particle is $m = 50$kDa, the mass moment of inertia $M = \frac{8}{15} m\sigma^2$ the translational and rotational diffusion constants, are $D_t = 1\mu\text{m}^2/\text{s}$ and $D_r = 1.6 \times 10^7 \text{rad}^2/\text{s}$ for all particles, the translational and rotational friction coefficients are $\gamma = \frac{k_\text{B}T}{D_t m}$ the $\Gamma = \frac{k_\text{B}T}{D_r M}$ respectively, where $k_\text{B} = 1.38 \times 10^{-23}\text{JK}^{-1}$ is the Boltzmann constant and $T = 300$K is the temperature of the system. To check if our Green's function and BD correctly describe the rotational motion of the particles, we have measured the correlation function of the dot product of the patch vector at a given time t with the initial patch vector. This correlation function decays as $e^{-2D_r t}$, where $D_r = k_\text{B}T/\Gamma M$, as expected.

In the following subsections we first present the results of the FFS-BD pre-simulation used to determine the value of the intrinsic dissociation rate k_d. Next, using the value of k_d, we perform MD-GFRD simulations in which we compute the probability that A and B are bound, as a function of system size. We compare the results against Monte Carlo simulations and analytical expressions. We compute the power spectra for the binding process. Finally, we discuss the performance of the algorithm.

3.3.1. RATE CONSTANT DETERMINATION USING FFS-BD PRE-SIMULATION

As explained in Sec. 3.2.6, it is advantageous to treat the dimer A-B as a single particle C, which then can dissociate again into A and B with an intrinsic rate k_d. We used direct FFS to precompute the intrinsic rate constant k_d. The interfaces λ_i are defined in terms of the interaction energy, as shown in Fig. 3.7. The bound state interface λ_{-1} was defined by $U(R,r) < -10k_\text{B}T$, the dissociated state final interface λ_5 was set at a distance $R = 1.6\sigma$. Five intermediate interface were set at respectively $\lambda_0 = -10k_\text{B}T$, $\lambda_1 = -2.5k_\text{B}T$, $\lambda_2 = -0.75k_\text{B}T$, $\lambda_3 = -0.025k_\text{B}T$, $\lambda_4 = -0.0075k_\text{B}T$. A straightforward BD trajectory created 100,000 configurations at the first interface. Subsequently, performing direct FFS yielded 20,000 configurations for each successive interface. Using Eq. 6.1, we find for the intrinsic dissociation rate constant $k_\text{d} = 4.66\text{s}^{-1}$. The configurations at the final interface can be used to draw from when performing the dissociation step in the MD-GFRD, see Sec. 3.2.5.

3.3.2. BIMOLECULAR REACTIONS

To test the multi-scale scheme, we simulate the bi-molecular reaction shown in Eq. 7.8. In these simulations we start off with two species of particles A and B, each having one patch on its surface. An A particle can react with a B particle to form a dimer. Also, a C particle can dissociate to form one A and one B, with an intrinsic rate k_d that has been pre-computed using FFS (see previous section). We assume that the mixture is ideal: only species A and B have an attractive interaction $U(R,r)$. All other interaction potentials between pairs A-A, B-B, C-C, C-A and C-B are repulsive only. We test the scheme for two different scenarios, one starting with a single A and a single B particle and the second starting with two A and two B particles. The simulation results are compared with Monte Carlo simulations of the same model and with analytical expressions.

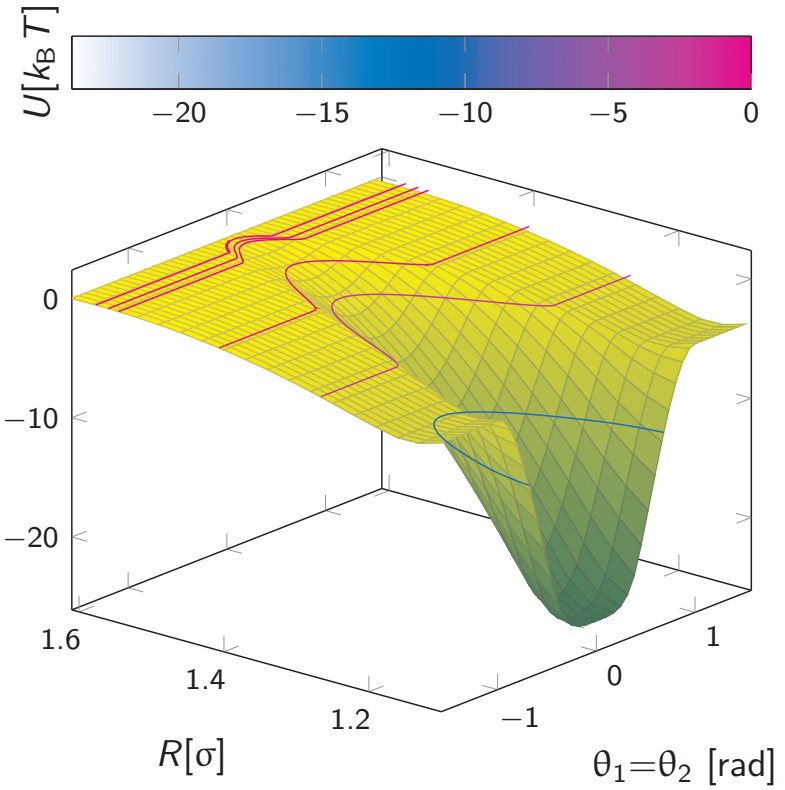


Figure 3.7: FFS interfaces were defined by the potential energy: $\lambda_0 = -10k_BT$, $\lambda_1 = -2.5k_BT$, $\lambda_2 = -0.75k_BT$, $\lambda_3 = -0.025k_BT$, $\lambda_4 = -0.0075k_BT$. The final interface λ_5 was defined by zero energy and a distance $R > 1.6\sigma$. Using these interfaces as starting points for successive trial runs, the particles are driven from the bound to the unbound state.

In the first case, one particle of species A and one particle of species B, each having one patch, are put in a cubic box of volume V, with periodic boundary conditions. This means that the number of C particles, N_C, is either zero or one. From the computed time average of N_C, we calculate the probability $\mathscr{P}_b$ that the A particle is bound to B. We repeat this procedure for different box sizes. In Fig. 3.8 we compare the value of $\mathscr{P}_b$ obtained using the new MD-GFRD algorithm to the results obtained from Monte Carlo simulations of the same system. The figure also shows the analytical result

$$\mathscr{P}_b = \frac{\langle N_C \rangle}{N_A} = \frac{k_{on}}{k_{on} + V k_{off}} = \frac{\phi(V)}{\phi(V)+1}, \tag{3.11}$$

where $\langle N_C \rangle$ is the average of N_C and $\phi(V)$ is the ratio of the probability that an A particle

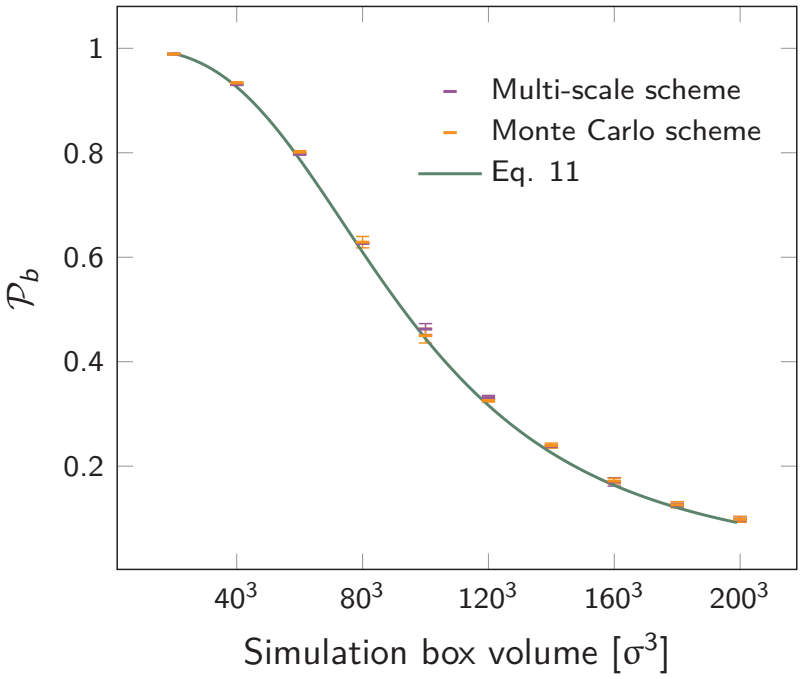


Figure 3.8: The probability $\mathscr{P}_b$ that a particle A is bound to a particle B, as a function of the volume of the box. Simulations are performed with one A particle and one B particle in the box. The points with the error bars are the results of the MD-GFRD simulations and the Monte Carlo simulation. These results are validated with the analytical prediction of Eq. 3.11. It is seen that the agreement is very good. The translational and rotational diffusion constants, which are not important for the value of $\mathscr{P}_b$, are $1\mu m^2/s$ and $1.6 \times 10^7 rad^2/s$.

is bound versus unbound

$$\phi(V) = \frac{k_{on}}{V k_{off}} = \frac{K_{eq}}{V}. \tag{3.12}$$

Here, k_{on} and k_{off} are the effective association and dissociation rates, respectively, and K_{eq} is the equilibrium constant

$$K_{eq} = \int d\mathbf{R} \int d\hat{\mathbf{u}}_1 \int d\hat{\mathbf{u}}_2 e^{-\beta V(R, r(\mathbf{R}, \hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2))}, \tag{3.13}$$

where $U(R, r(\mathbf{R}, \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_2))$ is the interaction potential given by Eq. 6.29, with $\mathbf{R}$ the inter-particle vector, R the magnitude of $\mathbf{R}$, r the distance between the patches of the particles, which depends on $\mathbf{R}$ and the orientation of the two particles denoted by the patch vectors in the stationary lab frame, $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$, respectively given by Eq. 3.1. Solving Eq. 3.13 analytically is not possible for the complex anisotropic potential used here. However, recently we have shown how in one TIS/FFS simulation both the association rate k_{on} and the dissociation rate k_{off} can be computed [23], which then allows us to obtain K_{eq}

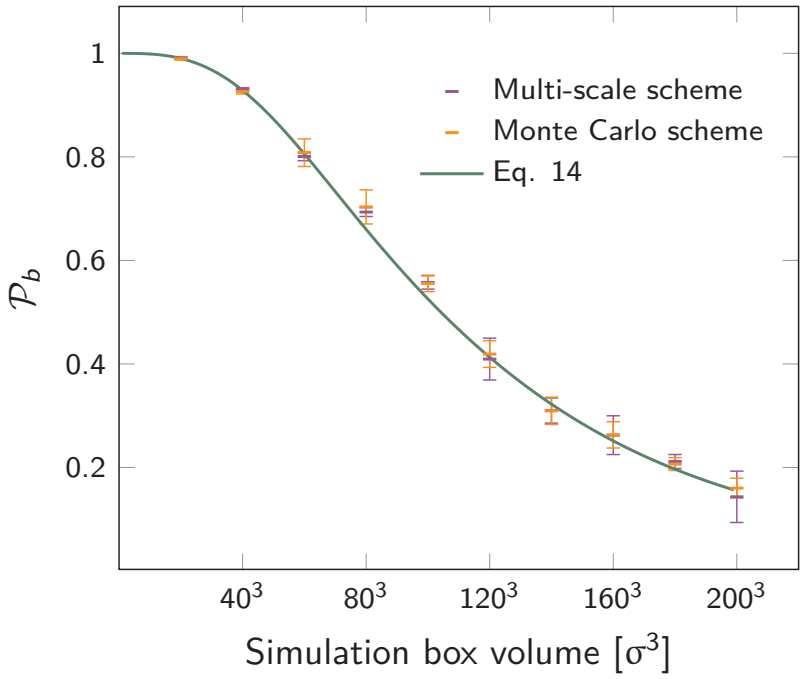


Figure 3.9: The probability $\mathscr{P}_b$ that a particle A is bound to a particle B, as a function of the volume of the box. Simulations are performed starting with two A particles and two B particles in the box. The points with the error bars show the results of the new MD-GFRD scheme and the Monte Carlo simulations. These results are validated with the analytical prediction of Eq. 3.14. It is seen that the agreement is very good. The translational and rotational diffusion constants, which are not important for the value of $\mathscr{P}_b$, are $1\mu m^2/s$ and $1.6\times10^7 rad^2/s$.

from Eq. 3.12. Applying this technique to this potential revealed that $k_{on} = 0.135\mu m^3 s^{-1}$ and $k_{off} = 1.384 s^{-1}$. Fig. 3.8 shows that the results of the MD-GFRD simulations agree very well with both the results of the Monte Carlo simulations and with the analytical predictions.

In the second test, we start with 2 A particles and 2 B particles, which can again interact via the same interaction potential to form species C. We can analytically compute the probability that an A particle is bound to a B particle, by carefully summing over all possible configurations [24]:

$$\mathscr{P}_b = \frac{\phi(V)+\phi(V)^2}{2(0.25+\phi(V)+\frac{\phi(V)^2}{2})}, \tag{3.14}$$

where $\phi(V)$ is given by Eq. 3.12. The results of the MD-GFRD simulations, the Monte Carlo simulations, and the analytical prediction are shown in Fig. 3.9. It is seen that the agreement is very good.

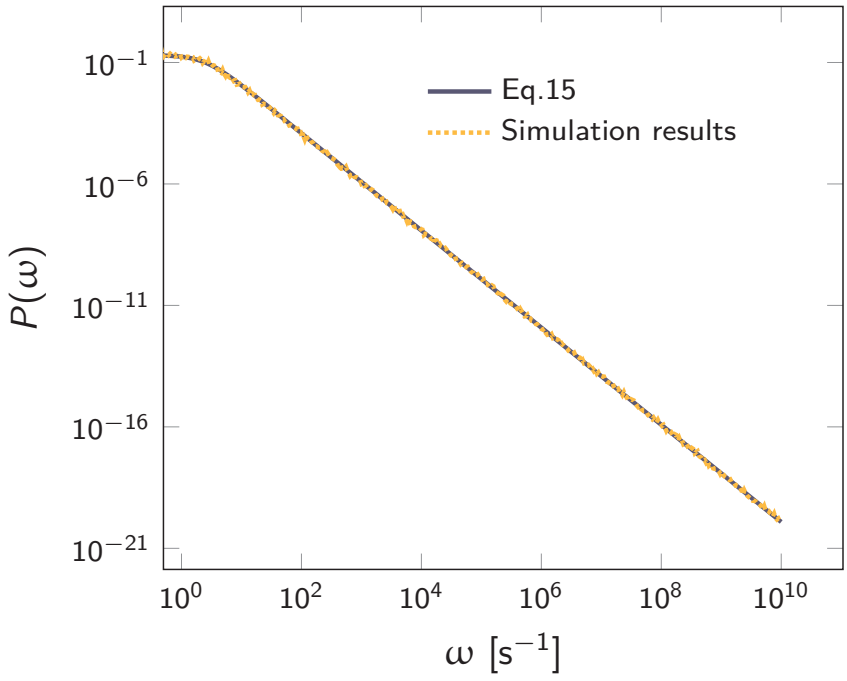


Figure 3.10: MD-GFRD successfully predicts the power spectrum $P_n(\omega)$ of the binding state $n(t)$ of two particles switching between the bound state $n(t) = 1$ and the unbound state $n(t) = 0$. The dotted line shows the results of the MD-GFRD simulations, while the solid line shows the analytical prediction of Eq. 3.15, where the association rate k_{on} and dissociation rate k_{off} have been computed from a single FFS simulation as described in Ref. [23]. Two particles, one of each species A and B were simulated in a box of side length 100σ.

3.3.3. POWER SPECTRUM

We can use MD-GFRD to compute the power spectrum $P_n(\omega)$ of the time trace of the binding state $n(t)$ of two particles, switching between the bound state with $n(t) = 1$ and the unbound state with $n(t) = 0$. The dotted line in Fig. 3.10 shows the result. We expect that this power spectrum is given by that of a random telegraph process [25]:

$$P(\omega) = \frac{2\mu\mathscr{P}_b(1-\mathscr{P}_b)}{\mu^2+\omega^2}, \tag{3.15}$$

where ω is the frequency, $\mu = k_{on}/V + k_{off}$ is the renormalized/effective decay rate, and $\mathscr{P}_b = k_{on}/(k_{on} + Vk_{off})$ is the binding probability. To predict the power spectrum, we thus need the effective association rate k_{on} and the effective dissociation rate k_{off}. As described in the previous section, these rates can be computed in a single TIS/FFS simulation [23]. Using the computed values of the rate constants in combination with Eq. 3.15, we arrive at the analytical prediction of the solid line in Fig. 3.10. It is seen that the

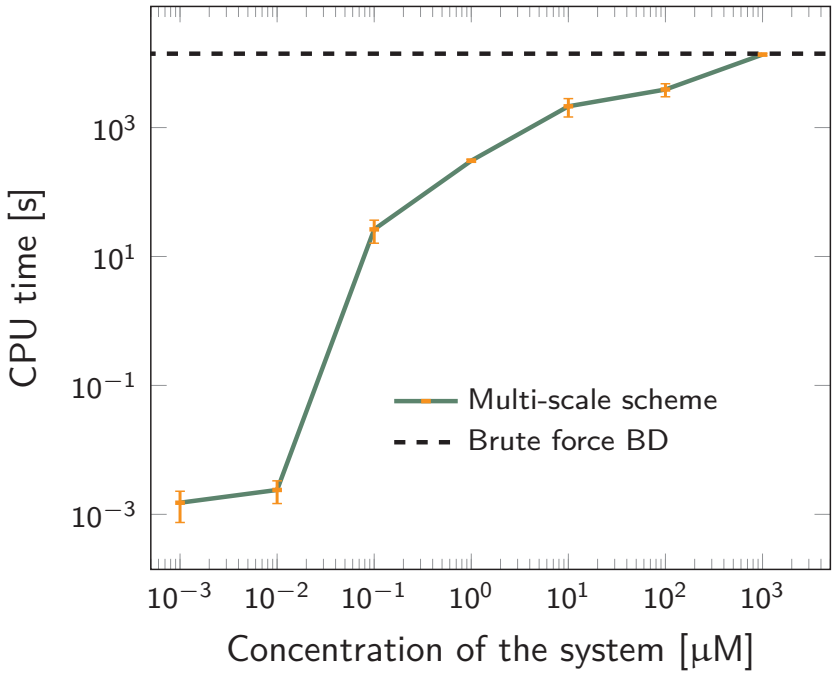


Figure 3.11: The CPU time to simulate 1ms real time as a function of the concentration of A and B, for MD-GFRD (solid line) and BD (dashed line). The concentration is varied by changing the volume of the simulation box, while the number of particles is kept constant at $N_A = N_B = 5$. It is seen that in the biologically relevant concentration range of nanomolar to micromolar the performance of MD-GFRD is much better than that of brute-force BD, but at higher concentrations the relative performance of MD-GFRD goes down. This is because at higher concentrations, the particles will be close to each other, and the system cannot capitalize on the potential of MD-GFRD to make large jumps in time and space.

agreement with the MD-GFRD simulation results is excellent. MD-GFRD thus not only reproduces mean quantities but also successfully predicts dynamic quantities.

3.3.4. PERFORMANCE

The motivation to combine GFRD and MD into a multi-scale scheme is the computational speed up it can provide. Unlike brute force Brownian dynamics which spends a lot of CPU time in propagating the particles toward each other, GFRD makes large jumps in space and time when the particles are far apart from each other and the GFRD domains are large. The computational power of GFRD can thus especially be reaped when the particles are often far apart, which is the case when the concentrations are low. This can be seen in Fig. 3.11, which shows a comparison of MD-GFRD against brute force BD as a function of concentration. It is seen that MD-GFRD is much more efficient than brute force BD, especially when the concentrations are below a μM. However, for high

concentrations, the performance of MD-GFRD becomes comparable to that of BD. In this regime, the particles are often so close together that no big jumps in time and space can be made. Interestingly, however, the crossover happens only at a mM concentration, which means that for most biologically relevant concentrations MD-GFRD is much faster than brute-force BD.

We performed a profiling of the code to establish the overhead associated with the GFRD checks. To determine the percentage of the total time the code spends in building domains and updating positions, the code was first profiled for a low concentration where all the particles are in GFRD. In that case, 50% of the time is spent on nearest neighbour searches and the other 50% is spent on constructing and bursting the domains and updating particles. To determine the overhead associated with GFRD compared to a brute force BD simulation, we profiled the code for a high concentration such that all the particles are in the BD regime. The code spends 60% of the total time integrating the BD particles (neighbour searches, force calculations, torque calculations and updating position and orientation) and 40% of the time attempting to build domains on these BD particles. The overhead associated with the nearest neighbour searches can be decreased, if the code is optimised such that the nearest neighbour search is performed just once for both the BD and GFRD.

3.4. CONCLUSION

In this work we extended the MD-GFRD scheme [11] to include the orientational dynamics of the particles, enabling the simulation of reaction and diffusion of particles that interact via anisotropic interaction potentials. This opens up the possibility to treat a whole class of interesting problems. Biomolecules such as proteins and DNA typically interact with each other via anisotropic potentials. In some cases of biological interest the dynamics at short length and time scales can be integrated out [26, 27, 28, 3]. For example, a gene regulatory protein that has just dissociated from its promoter on the DNA either rapidly rebinds the DNA or rapidly escapes into the bulk, where it will loose its orientation; conversely, a new protein tends to arrive at the promoter from the bulk in a random orientation. In these cases, we expect that the regulatory proteins can be modeled as isotropic particles that interact with the DNA via effective rate constants, which take into account the anisotropy of the interaction. However, it is now well established that in many systems the dynamics at molecular length and timescales, arising from e.g. enzyme-substrate rebindings, can qualitatively change the macroscopic behavior of the system at cellular length scales [10, 29]. This phenomenon can occur in biochemical networks with multi-site protein modification, which are omnipresent in cellular biology [10]. In such systems, the orientational dynamics cannot be integrated out: the probability that an enzyme which has dissociated from its substrate molecule rebinds to another site on the same substrate molecule to chemically modify it, will depend in a non-trivial manner on the translational and orientational diffusion constants of the particles, their size, and the distance between the patches on the substrate. The MD-GFRD scheme presented here now makes it possible to study the interplay between the microscopic dynamics at the molecular scale and the network dynamics at the cellular scale in this large class of systems.

In addition, the MD-GFRD scheme could more generally be used for soft matter self-assembly where building blocks that are diffusing in the dilute solution come together and bind occasionally to form large complexes and structures[14].

While the MD-GFRD scheme has been set up for simulating 3D bulk solutions, in principle the method can be extended to include other geometries, where GFRD includes reaction and diffusion in 1D and 2D [30, 31, 32].

Finally, we note that in our multiscale MD-GFRD algorithm we assume that the interaction potentials are short ranged, which is usually the case in a highly screened environment such as the living cell. Moreover, the algorithm, similar to most algorithms to simulate biochemical networks in time and space, assumes that the particles move by normal diffusion [33, 4, 5, 7, 34], as indeed experiments suggest proteins do inside the living cell [1]. The algorithm cannot straightforwardly treat long-range hydrodynamics interactions, which can be of importance under certain (non-equilibrium) conditions. For future work it would be of interest to study whether hydrodynamic interactions could be included in MD-GFRD, for instance via the Oseen tensor approximation.

3.5. ACKNOWLEDGEMENTS

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.

BIBLIOGRAPHY

[1] Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–1194.

[2] Schreiber G, Haran G, Zhou HX (2009) Fundamental Aspects of ProteinProtein Association Kinetics. *Chemical Reviews* 109:839–860.

[3] ten Wolde PR, Becker NB, Ouldridge TE, Mugler A (2016) Fundamental Limits to Cellular SensingPieter Rein ten Wolde, Nils B. Becker, Thomas E. Ouldridge & Andrew Mugler. *Journal of Statistical Physics* 162:1395–1424.

[4] Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Bio.* 1:137–151.

[5] Lipková J, Zygalakis KC, Chapman SJ, Erban R (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. *J. Appl. Math.* 71:714–730.

[6] Flegg MB, Chapman SJ, Erban R (2012) The two-regime method for optimizing stochastic reaction-diffusion simulations. *J. R. Soc. Interface* 9:859–868.

[7] Schöneberg J, Noé F (2013) ReaDDy - A Software for Particle-Based Reaction-Diffusion Dynamics in Crowded Cellular Environments. *PLoS ONE* 8:e74261.

[8] van Zon JS, ten Wolde PR (2005) Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Physical Review Letters* 94:128103.

[9] van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.* 123:234910.

[10] Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *P. Natl. Acad. Sci. USA* 107:2473–2478.

[11] Vijaykumar A, Bolhuis PG, ten Wolde PR (2015) Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations. *J. Chem. Phys.* 143:214102.

[12] van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762.

[13] Allen RJ, Warren PB, ten Wolde PR (2005) Sampling Rare Switching Events in Biochemical Networks. *Phys. Rev. Lett.* 94:018104.

[14] Glotzer SC, Solomon MJ (2007) Anisotropy of building blocks and their assembly into complex structures. *Nat. Mater.* 6:557–562.

[15] Newton AC, Groenewold J, Kegel WK, Bolhuis PG (2015) Rotational diffusion affects the dynamical self-assembly pathways of patchy particles. *Proc. Nat. Acad. Sci. USA* 112:15308–15313.

[16] Davidchack R, Ouldridge T, Tretyakov M (2015) New langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* 142.

[17] Loman A, Gregor I, Stutz C, Mund M, Enderlein J (2010) Measuring rotational diffusion of macromolecules by fluorescence correlation spectroscopy. *Photochem. Photobiol. Sci.* 9:627–636.

[18] Versmold H (1970) Time correlation function for internal and anisotropic rotational motion of molecules. *Z. Naturforsch* 25:367–372.

[19] Favro LD (1960) Theory of the rotational Brownian motion of a free rigid body. *Phys. Rev.* 119:55–62.

[20] Furry WH (1957) Isotropic rotational Brownian motion. *Phys. Rev.* 107:7–.

[21] Allen RJ, Valeriani C, Rein ten Wolde P (2009) Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter* 21:463102.

[22] Ouldridge TE, Louis AA, Doye JPK (2011) Structural, mechanical and thermodynamic properties of a coarse-grained model of DNA. *J. Chem. Phys.* 134:085101.

[23] Vijaykumar A, Bolhuis PG, ten Wolde PR (2016) The intrinsic rate constants in diffusion-influenced reactions. *Faraday Discuss.* in press.

[24] Ouldridge TE, Louis AA, Doye JPK (2010) Extracting bulk properties of self-assembling systems from small simulations. *Journal of Physics: Condensed Matter* 22:104102.

[25] Kaizu K, et al. (2014) The Berg-Purcell Limit Revisited. *Biophysical Journal* 106:976–985.

[26] van Zon JS, Morelli MJ, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophysical Journal* 91:4350–4367.

[27] Morelli MJ, Allen RJ, ten Wolde PR (2011) Effects of Macromolecular Crowding on Genetic Networks. *Biophysical Journal* 101:2882–2891.

[28] Mugler A, ten Wolde PR (2013) The Macroscopic Effects of Microscopic Heterogeneity in Cell Signaling. *Advances in Chemical Physics, Volume 153* pp 373–396.

[29] Aokia K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:12675–12680.

[30] Paijmans J, ten Wolde PR (2014) Lower bound on the precision of transcriptional regulation and why facilitated diffusion can reduce noise in gene expression. *Phys. Rev. E* 90:032708.

[31] Wehrens M, ten Wolde PR, Mugler A (2014) Positive feedback can lead to dynamic nanometer-scale clustering on cell membranes. *J. Chem. Phys* 141:205102.

[32] Sokolowski TR (2014) Ph.D. thesis (VU University Amsterdam).

[33] Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst Biol (Stevenage)* 1:230–236.

[34] Johnson ME, Hummer G (2014) Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. *Phys. Rev. X* 4:031037.

4

THE INTRINSIC RATE CONSTANTS IN DIFFUSION-INFLUENCED REACTIONS

Intrinsic rate constants play a dominant role in the theory of diffusion-influenced reactions, but usually as abstract quantities that are implicitly assumed to be known. However, recently it has become clear that modelling complex processes requires explicit knowledge of these intrinsic rates. In this paper we provide microscopic expressions for the intrinsic rate constants for association and dissociation processes of isotropically interacting particles and illustrate how these rates can be computed using rare event simulations techniques. In addition, we address the role of the orientational dynamics, for particles interacting via anisotropic potentials.

4.1. INTRODUCTION

The association and dissociation of two particles are elementary steps in many processes in biology, such as receptor-ligand and enzyme-substrate binding in cell signalling, and protein-DNA binding in gene regulation. Also in materials science association and dissociation play a central role, e.g. in the self-assembly of colloidal particles, the formation of micro-emulsions, or the phase behavior of polymer solutions. In these processes, the particles typically come into contact via diffusion, after which they bind with a rate that depends on the intrinsic association rate constant; conversely, the associated particles dissociate with an intrinsic dissociation rate, after which they move apart via diffusion.

In the past decades, theories of diffusion-influenced reactions have been developed that show how the effective rate constants depend on the diffusion constants of the particles, their cross section, their interaction potential, and the intrinsic association and dissociation rate constants [1]. However, these theories assume *a priori* given intrinsic association and dissociation rate constants. Similarly, techniques to simulate networks of chemical reactions have been developed, in which the particles typically have an idealized shape, move by diffusion, and react upon contact with *given* intrinsic rate constants [2, 3, 4, 5, 6, 7]. In parallel, simulation techniques have been developed that enable the calculation of association and dissociation rate constants for pairs of particles that interact via potentially complex interaction potentials [8, 9]. Yet, these techniques typically compute *effective* rate constants, which result from the combined effects of diffusion, the interaction potential, and binding upon contact. Moreover, also in experiments typically the effective rate constants are measured. How the intrinsic rate constants depend on the interaction potential, the cross section, and the diffusion constants of the particles, has thus received little attention.

In general, an association-dissociation reaction is a complicated non-Markovian many-body problem that cannot be solved analytically. The reason is that the process of binding generates non-trivial spatio-temporal correlations between the positions of the reactants, which depend on the history of the association and dissociation events. Capturing these correlations requires knowledge of not only the diffusion constants, the interaction potential, and the cross section, but also the intrinsic rate constants.

In dilute systems, however, typically only the effective rate constants are needed to describe the system's dynamics at long times [10, 11]. When the concentrations are low, the time it takes for two reactants to meet each other is much longer than the time they spend in close proximity: once the reactants are near each other, they either rapidly bind or rapidly diffuse back into the bulk. Similarly, after a dissociation event, the two reactants either quickly rebind or rapidly move away from each other. Under these conditions, it is often possible to integrate out the dynamics at the molecular scale, and describe the association-dissociation reaction as a Markovian process with effective association and dissociation rates [10, 11]. While these effective rates depend on the intrinsic rate constants, only the effective rate constants determine the dynamics at the relevant length and time scales [10, 11].

Yet, is now clear that even in dilute systems, spatio-temporal correlations at the molecular scale can dramatically change the behavior of the system at the macroscopic scale [4, 12]. In the case of multi-site protein modification, enzyme-substrate rebindings can lead to the loss of ultra-sensitivity and even bi-stability, essentially because rebindings

can turn a distributive mechanism into a processive one [4, 12]. In such a scenario, even the long-time dynamics cannot be described by effective rate constants: one needs to know the diffusion constants, the cross sections, and the intrinsic rate constants. Moreover, while spatial heterogeneity at the molecular scale can arise from these non-trivial spatio-temporal correlations, in cellular systems microscopic heterogeneity is also imposed via molecular structures. In fact, it is now becoming increasingly recognized that cells exploit the spatial heterogeneity of micro-domains, lipid rafts, clusters, and scaffolds as a computational degree of freedom for enhancing information transmission [13, 14]. Modeling the reactions in these spatially heterogeneous systems often requires knowledge of the intrinsic rate constants. Last but not least, for simulating association and dissociation reactions in 1D and 2D, knowledge of the intrinsic rate constants is even more pertinent, because no well-defined effective rate constant exists in the long-time limit.

In this manuscript, we provide microscopic expressions for the intrinsic rate constants, and illustrate how these expressions can be used to compute rate constants in rare-event simulation techniques such as Transition Interface Sampling (TIS) [15, 16, 17] and Forward Flux Sampling (FFS) [18, 19]. While computing both the forward and backward rate typically requires two separate simulations, we will show how, by exploiting analytical expressions for the binding and escape probability, both the association and dissociation rate constants (and hence the equilibrium constant) can be obtained in one single simulation. We discuss the relationship with the technique developed by Northrup and coworkers [8] for computing effective association rates. Finally, we address the role of orientational dynamics in association and dissociation reactions.

4.2. THEORY OF DIFFUSION-INFLUENCED REACTIONS

We consider two particles A and B that interact via an isotropic interaction potential $U(r)$, and move with an interparticle diffusion constant $D = D_A + D_B$, where D_A and D_B are the diffusion constants of A and B, respectively. Upon contact at the interparticle distance σ, the particles can associate with a rate that is determined by the intrinsic associate rate constant k_a, and, when bound, the two can dissociate with an intrinsic dissociation rate k_d. Following the work of Agmon and Szabo [1], we rederive in the appendix the following central results. The effective association rate constant is given by

$$k_{\mathrm{on}} = \frac{k_a p_{\mathrm{eq}}(\sigma) k_D}{k_a p_{\mathrm{eq}}(\sigma) + k_D} = [1 - S_{\mathrm{rad}}(t \to \infty|\sigma)]\, k_D. \tag{4.1}$$

Here, k_D is the diffusion-limited rate constant, which determines the rate at which the two particles diffuse towards each other, and $p_{\mathrm{eq}}(\sigma) \simeq e^{-\beta U(\sigma)}$, with $\beta = 1/(k_B T)$ the inverse temperature, is the equilibrium probability that they are at the distance σ. The survival probability $S_{\mathrm{rad}}(t \to \infty|\sigma) = k_D/(k_a p_{\mathrm{eq}}(\sigma) + k_D)$ is the probability that the two particles, given that they are at contact, escape into the bulk before binding to each other. Hence the effective association rate is given by the rate at which the two particles get in contact, which is determined by k_D, and the probability $1 - S_{\mathrm{rad}}(t \to \infty|\sigma) = k_a p_{\mathrm{eq}}(\sigma)/(k_a p_{\mathrm{eq}}(\sigma) + k_D)$ that upon contact, they bind.

The effective dissociation rate is given by

$$k_{\text{off}} = \frac{k_d k_D}{k_a p_{\text{eq}}(\sigma) + k_D} = k_d S_{\text{rad}}(t \to \infty | \sigma). \tag{4.2}$$

The effective dissociation rate is thus given by the dissociation rate k_d times the probability $S_{\text{rad}}(t \to \infty|\sigma) = k_D/(k_a p_{\text{eq}}(\sigma) + k_D)$ that the particles upon dissociation diffuse away from each other and escape into the bulk.

It can also be verified that the equilibrium constant is given by

$$K_{\text{eq}} = \frac{k_a p_{\text{eq}}(\sigma)}{k_d} = \frac{k_{\text{on}}}{k_{\text{off}}}. \tag{4.3}$$

At this stage, a few points are worthy of note. First, these results hold for isotropic, but otherwise arbitrary interaction potentials $U(r)$. Secondly, the contact distance σ serves to define the dividing surface that separates the bound from the unbound state. This surface is usually taken to be near the free-energy barrier that separates the bound from the unbound state. The precise location of this dividing surface is somewhat arbitrary, as the effective rate constants can by definition not depend on the choice made for σ. This is in marked contrast to the intrinsic rate constants k_a and k_d and the diffusion-limited rate k_D, which all sensitively depend on σ. Thirdly, the diffusion-limited rate constant depends not only on σ and D, but also on the interaction potential $U(r)$. For arbitrary interaction potentials $U(r)$, no analytical expression for k_D is, in general, available. However, when σ is chosen to be beyond the range r_c of the interaction potential, then an exact expression is well known—the Smoluchowski diffusion-limited reaction rate constant [20]:

$$k_D = 4\pi\sigma D. \tag{4.4}$$

Moreover, when $\sigma > r_c$, then $U(\sigma) = 0$, and $p_{\text{eq}}(\sigma) = 1$ In this scenario, the effective rate constants are given by:

$$k_{\text{on}} = \frac{k_a(\sigma) k_D(\sigma)}{k_a(\sigma) + k_D(\sigma)} \tag{4.5}$$

$$k_{\text{off}} = \frac{k_d(\sigma) k_D(\sigma)}{k_a(\sigma) + k_D(\sigma)}. \tag{4.6}$$

Here, and below, we have written $k_a = k_a(\sigma)$, $k_d = k_d(\sigma)$ and $k_D = k_D(\sigma)$ to remind ourselves that these rate constants, in contrast to the effective rate constants k_{on} and k_{off}, depend on our choice for σ.

4.3. EFFECTIVE POSITIVE FLUX EXPRESSION

To obtain the intrinsic rate constants in computer simulations, we need expressions in terms of microscopic quantities that can be measured in the simulation. We will focus on the dissociation pathway, and the dissociation rate k_{off}. To compute this rate, we use the "effective positive flux" expression of van Erp and coworkers [15, 16, 17]. The progress of the dissociation reaction is quantified via a parameter $\lambda(r)$, which depends

on the separation r between the particles A and B. For simplicity, we use $\lambda = r$ A series of interfaces is chosen, $r_0, r_1, \ldots, r_{n-1}, r_n$, such that r_0 is deep in the bound state and r_n is in the unbound state. Strictly speaking the unbound state is defined by $r_n \to \infty$, a point to which we will return in the next section. Defining the history-dependent functions indicator $h_{\mathscr{B}}$ and $h_{\mathscr{U}}$ such that $h_{\mathscr{B}} = 1$ and $h_{\mathscr{U}} = 0$ if the system was more recently in the bound state B ($r < r_0$) than in the unbound state U ($r > r_n$), and $h_{\mathscr{B}} = 0$ and $h_{\mathscr{U}} = 1$ otherwise, the rate constant k_{off} for transitions from B to U is given by [15]

$$k_{\text{off}} = \frac{\overline{\Phi}_{B,n}}{\overline{h}_{\mathscr{B}}} = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} P(r_n|r_0). \tag{4.7}$$

Here, $\Phi_{B,j}$ is the flux of trajectories coming from the bound state B (with $r < r_0$) that cross r_j for the first time; thus, $\Phi_{B,n}$ is the flux of trajectories from the bound state towards the unbound state, $r > r_n$, and $\Phi_{B,0}$ is the flux reaching the first interface r_0. The factor $\overline{h}_{\mathscr{B}}$ is the average fraction of time that the system spends in the bound state B. $P(r_n|r_0)$ is the probability that a trajectory that reaches r_0 subsequently arrives at interface r_n instead of returning to the bound state r_0. The expression thus states that the total flux of trajectories from the bound state to the unbound state is the flux of trajectories from B to r_0 multiplied by the probability that such a trajectory will later reach r_n before returning to r_0. $P(r_n|r_0)$ can be expressed as the product of the probabilities $P(r_{i+1}|r_i)$ that a trajectory that comes from r_0 and crosses r_i for the first time will subsequently reach r_{i+1} instead of returning to r_0:

$$P(r_n|r_0) = \prod_{i=0}^{n-1} P(r_{i+1}|r_i). \tag{4.8}$$

Combining Eqs. 4.7 and 4.8, the effective dissociation rate can thus be expressed as

$$k_{\text{off}} = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} \prod_{i=0}^{n-1} P(r_{i+1}|r_i). \tag{4.9}$$

The individual factors $P(r_{i+1}|r_i)$ can be determined in a Transition Interface Sampling [15] (TIS) or a Forward Flux Sampling [18, 19] (FFS) simulation, as the fraction of trajectories crossing the interface r_i that reach the interface r_{i+1} instead of returning to r_0. FFS and TIS are both based on the effective positive flux expression, Eq. 4.7, but differ in the way they construct the path ensembles.

4.4. INTRINSIC DISSOCIATION RATE AND EFFECTIVE POSITIVE FLUX

To obtain a microscopic expression for the intrinsic dissociation rate k_d, we rewrite Eq. 4.9 as

$$k_{\text{off}} = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} P(r_{n'}|r_0) P(r_n|r_{n'}), \tag{4.10}$$

where $P(r'_n|r_0) = \prod_{i=0}^{n'-1} P(r_{i+1}|r_i)$ and $P(r_n|r_{n'}) = \prod_{i=n'}^{n-1} P(r_{i+1}|r_i)$. Now, the crux is to define $r_{n'} = \sigma$. Comparing Eq. 4.10 with Eq. 4.2, we can then make the following identifications:

$$k_d(\sigma) = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} P(\sigma|r_0) \tag{4.11}$$

$$S_{\text{rad}}(t \to \infty|\sigma) = \frac{k_D(\sigma)}{k_a p_{\text{eq}}(\sigma) + k_D(\sigma)} = P(r_n|\sigma) \tag{4.12}$$

Eq. 4.11 provides a microscopic expression for the intrinsic dissociation rate, and is one of the central results of this paper. It shows that the intrinsic dissociation rate is the flux of trajectories that come from the bound state and cross the dividing surface $r_{n'} = \sigma$. The expression makes explicit that the intrinsic dissociation rate depends on the choice for σ. Also Eq. 4.12 highlights the idea that not only the intrinsic rates, but also the diffusion-limited rate k_D depends on this choice. We further iterate that σ need not be chosen beyond the interaction range of the potential; the expressions hold for any choice of σ.

The microscopic expressions of Eqs. 4.11 and 4.12 make it possible to obtain both the effective dissociation rate k_{off} and the intrinsic dissociation rate k_d from computer simulations. Again, Transition Interface Sampling [15] and Forward Flux Sampling [18, 19] are particularly well suited, because they are both based on the effective positive flux expression, Eq. 4.7.

In fact, by choosing the cross-section σ beyond the range r_c of the interaction potential, it is possible to obtain from one simulation not only the intrinsic dissociation rate k_d and effective dissociation rate k_{off}, but also the intrinsic association rate k_a and effective association rate k_{on}, and hence the equilibrium constant K_{eq}. One TIS/FFS simulation yields both k_d, from Eq. 4.11, and $P(r_n|\sigma) = S_{\text{rad}}(t \to \infty|\sigma)$, from Eq. 4.12. Yet, when $\sigma > r_c$ (and $U(\sigma) = 0$), we know that the latter is also given by

$$P(r_n|\sigma) = S_{\text{rad}}(t \to \infty|\sigma) = \frac{k_D(\sigma)}{k_a(\sigma) + k_D(\sigma)}, \tag{4.13}$$

with $k_D = 4\pi\sigma D$, as discussed in section 4.2. In other words, having computed $P(r_n|\sigma)$ in the TIS/FFS simulation, we can use the above expression and the analytical solution $k_D = 4\pi\sigma D$, to obtain not only k_d but also k_a:

$$k_a(\sigma) = \frac{(1 - P(r_n|\sigma))k_D(\sigma)}{P(r_n|\sigma)}. \tag{4.14}$$

From k_d and k_a, we obtain the equilibrium constant $K_{\text{eq}} = k_a/k_d$, from which we then find $k_{\text{on}} = K_{\text{eq}} k_{\text{off}}$.

As pointed out above, the effective rates are strictly defined for $r_n \to \infty$, meaning that Eqs. 4.11 and 4.14 are also only valid in that limit. However, to keep the simulations tractable, in practice one would like to use an interface at finite r_n. In the next section we derive an expression that holds for finite interface values r_n.

4.5. COMPUTING THE INTRINSIC RATES FOR AN INTERFACE AT FINITE R_n

While the intrinsic association rate constant k_d does not rely on taking the position of the last interface $r_n \to \infty$, the intrinsic rate constant k_a does (see Eq. 4.14). To obtain an expression for k_a for a *finite* value of r_n, we start by rewriting the effective association rate Eq. 4.6 as the sum of reciprocal intrinsic rates:

$$\frac{1}{k_{\text{on}}} = \frac{1}{k_a(\sigma)} + \frac{1}{k_D(\sigma)}. \tag{4.15}$$

Since the effective association rates are independent on the choice of the dividing surface σ, we can choose to set the dividing surface at $r_n > \sigma$

$$\frac{1}{k_{\text{on}}} = \frac{1}{k_a(r_n)} + \frac{1}{k_D(r_n)}. \tag{4.16}$$

Combining the above two equations and rearranging yields

$$\frac{1}{k_a(r_n)} - \frac{1}{k_a(\sigma)} = \frac{1}{k_D(\sigma)} - \frac{1}{k_D(r_n)}. \tag{4.17}$$

To make progress we need to relate the intrinsic rate constants $k_a(\sigma)$ and $k_a(r_n)$. This relation is provided by linking the intrinsic dissociation rates, Eq. 4.11 at the respective surfaces, yielding

$$k_d(r_n) = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} P(r_n|r_0) = \frac{\overline{\Phi}_{B,0}}{\overline{h}_{\mathscr{B}}} P(\sigma|r_0) P(r_n|\sigma) = k_d(\sigma) P(r_n|\sigma). \tag{4.18}$$

Since detailed balance implies $K_{\text{eq}} = k_a(\sigma)/k_d(\sigma) = k_a(r_n)/k_d(r_n)$, the desired relation for the intrinsic association rate constants is

$$k_a(r_n) = k_a(\sigma) P(r_n|\sigma). \tag{4.19}$$

Inserting Eq. 4.19 into Eq. 4.17 and rearranging yields

$$k_a(\sigma) = \frac{(1 - P(r_n|\sigma))\, k_D(\sigma)}{P(r_n|\sigma)(1-\Omega)}, \tag{4.20}$$

where $\Omega \equiv k_D(\sigma)/k_D(r_n)$, which using $k_D(b) = 4\pi\sigma b$, reduces to $\Omega = \sigma/r_n$. Eq. 4.20 provides an explicit expression for the intrinsic association rate k_a for finite r_n, featuring a correction factor $1/(1-\Omega)$. For $r_n \to \infty$, Ω vanishes and the expression reduces to Eq. 4.14. However, the correction factor decays slowly with r_n, and in practice it cannot be neglected. Since Ω is known analytically, Eq. 4.20 turns this approach into a feasible strategy for computing both k_d and k_a in TIS and FFS, which directly give access to $P(r_n|\sigma)$. Lastly, combining this expression with Eqs. 4.5 and 4.6, we obtain the following expressions for the effective on and off rates, respectively:

$$k_{\text{on}} = \frac{(1 - P(r_n|\sigma))\, k_D(\sigma)}{1 - P(r_n|\sigma)\Omega}, \tag{4.21}$$

$$k_{\text{off}} = k_d(\sigma) P(r_n|\sigma) \frac{1-\Omega}{1 - P(r_n|\sigma)\Omega}. \tag{4.22}$$

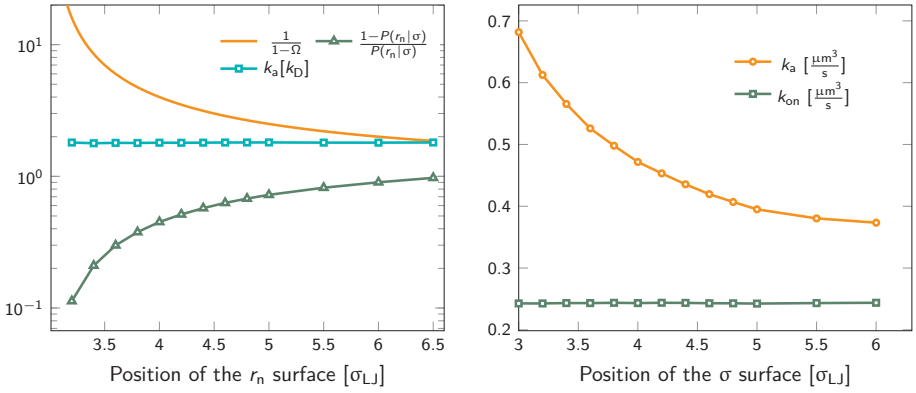


Figure 4.1: The intrinsic rate constant for the isotropic Lennard-Jones potential. Left: $k_a(\sigma)$ as a function of the position of last interface, r_n in units of $k_D(\sigma)$, for a fixed cross section $\sigma = 3\sigma_{LJ}$. The correction factor $\frac{1}{1-\Omega}$ and the factor $(1-P(r_n|\sigma))/P(r_n|\sigma)$ are also shown. As expected the value of $k_a(\sigma)$ remains constant as r_n is varied. Right: $k_a(\sigma)$ and k_{on} as a function of the cross section σ. The position of the last interface, r_n, is kept constant at $6.5\sigma_{LJ}$. It is seen that while k_a varies with σ, k_{on} does not.

4.5.1. ILLUSTRATIVE EXAMPLE

As an illustration of the above scheme we numerically evaluated Eq. 4.20 for a two particle system undergoing Brownian dynamics (BD). The interaction potential between the two particles is given by

$$u(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma_{LJ}}{r}\right)^{12} - \left(\frac{\sigma_{LJ}}{r}\right)^{6}\right], & 0 < r < 3\sigma_{LJ} \\ 0, & r > 3\sigma_{LJ}, \end{cases} \tag{4.23}$$

where σ_{LJ} sets the length-scale of the Lennard-Jones potential, ϵ sets the well depth, and r is the interparticle distance. In the simulations $\sigma_{LJ} = 5$nm, $\epsilon = 10 k_B T$. The length of the simulation box is $60\sigma_{LJ}$. We evaluate $k_d(\sigma)$ and $P(r_n|\sigma)$ using one single FFS simulation. First, we fix the cross section $\sigma = 3\sigma_{LJ}$, just beyond the cut-off distance of the potential. Interfaces were set at r={1.3, 1.5, 2.0, 2.5, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.5, 6.0, 6.5 }, in units of σ_{LJ}. From each interface, 10000 trajectories are started and the conditional probability as in Eq. 4.8 is calculated.

We test the independence of Eq. 4.20 on the location of the final interface by varying r_n, while keeping $\sigma = 3.0\sigma_{LJ}$ fixed. As seen in Fig. 4.1(left) the value of $k_a(\sigma)$ is independent of the position of the r_n surface. Note, however, that the correction factor $1/(1-\Omega)$ is not negligible, even for $r_n = 6.5\sigma_{LJ}$.

Next, we plot in Fig. 4.1(right) the dependence of the intrinsic association rate $k_a(\sigma)$ on the location of the cross section σ, keeping $r_n = 6.5\sigma_{LJ}$ constant. The intrinsic association rate constant decreases with σ, but the effective rate constant k_{on} is independent of σ, as expected.

The intrinsic dissociation rate $k_d(\sigma)$ is evaluated via Eq. 4.11. From $k_d(\sigma)$ and $k_a(\sigma)$,

the values of k_{on} and k_{off}, are evaluated using Eq. 4.5 and Eq. 4.6. K_{eq} is calculated using Eq. 4.3. The results are summarized in Table 4.1.

4.5.2. VALIDATION OF THE EFFECTIVE RATE CONSTANTS

To validate the values of k_{on} and k_{off} obtained from the FFS simulations, we calculate these values also from a brute force Brownian Dynamics simulation of the same system of two particles. First, the equilibrium constant K_{eq} is evaluated using the analytic expression

$$K_{eq} = 4\pi \int_0^{3\sigma_{LJ}} r^2 e^{-\beta u(r)} dr = \frac{k_{on}}{k_{off}}, \tag{4.24}$$

where $u(r)$ is given by Eq. 4.23, and the integral is over the bound state $0 < r < 3\sigma_{LJ}$.

The brute force simulation generates a time trace $n(t)$, switching from the bound state with $n(t) = 1$ to the unbound state with $n(t) = 0$. From this time trace we generate a time autocorrelation function

$$c(\tau) = \frac{\langle n(\tau) n(0) \rangle}{\langle n(0)^2 \rangle} \approx e^{-\mu\tau}, \tag{4.25}$$

that relaxes exponentially with a decay constant μ given by

$$\mu = k_{on}/V + k_{off}. \tag{4.26}$$

Fig. 4.2 shows this time autocorrelation function. The simulation results are fitted to Eq. 4.25 to obtain the value of μ and by combining with Eqs. 4.24 and 4.26, the effective rate constants are obtained.

Alternatively, we can generate a power spectrum $P(\omega)$ from the same $n(t)$ time trace [10]. The power spectrum for a random telegraph process with switching rates k_f and k_b is given by

$$P(\omega) = \frac{2\mu p(1-p)}{\mu^2 + \omega^2}, \tag{4.27}$$

where ω is the frequency, $\mu = k_f + k_b$ is the decay rate and $p = k_f/(k_f + k_b)$. The low-frequency part of the power spectrum as obtained from the simulations is expected to be given by the above expression, with $k_f = k_{on}/V$ and $k_b = k_{off}$ (see Ref. [10]). Table 1 compares the values of k_{on} and k_{off} as obtained via this scheme, with those from FFS, using Eq. 4.3 and Eqs. 4.18–4.22, and the results from the time auto-correlation function, Eqs. 4.25 and 4.26. It is seen that the values are in very good agreement.

	$K_{eq}[10^{-3}\mu m^3]$	$k_{on}[\frac{\mu m^3}{s}]$	$k_{off}[\frac{1}{s}]$
FFS	5.127	0.2417	47.14
Time autocorrelation	5.145	0.2589	50.33
Power Spectrum	5.145	0.2571	49.99

Table 4.1: Comparison of the effective rate constants as determined via different approaches: FFS, using Eq. 4.3 and Eqs. 4.18–4.22; the autocorrelation function, Eqs. 4.25 and 4.26; the power spectrum, Eq. 4.27. The table shows good agreement between the values of k_{on} and k_{off}, and hence K_{eq}, obtained via these different approaches.

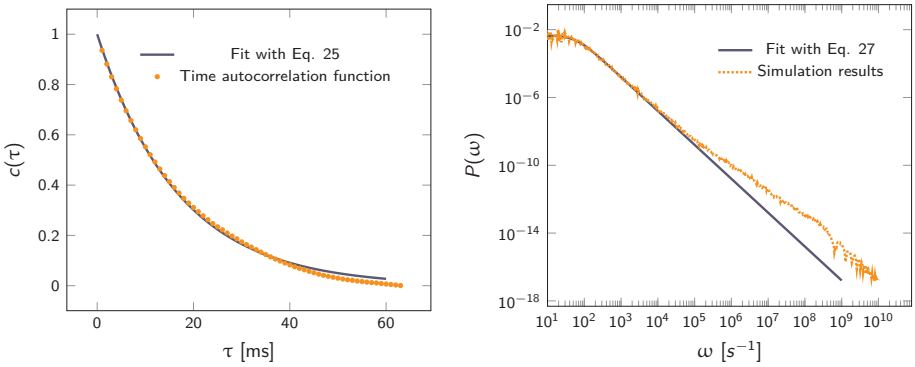


Figure 4.2: Left: The autocorrelation function $c(\tau)$ obtained from simulation is fit to Eq. 4.25, yielding $\mu = 59.9236\text{s}^{-1}$. Right: Power spectrum $P(\omega)$ fitted to Eq. 4.27 to determine the value of $\mu = 59.5192\text{s}^{-1}$. In both cases the effective rate constants are calculated by solving Eq. 4.26 and Eq. 4.24.

4.6. RELATION TO OTHER TECHNIQUES

Northrup et al. [8, 21] provide a method to compute the effective association rate directly from Eq. 4.1

$$k_{\text{on}} = [1 - S_{\text{rad}}(t \to \infty|\sigma)]\, k_D(\sigma) \equiv \beta_\infty k_D(\sigma), \tag{4.28}$$

where $\beta_\infty = 1 - S_{\text{rad}}(t \to \infty|\sigma)$ is defined as the probability that particles starting at a distance σ associate rather than diffuse away and escape into the bulk. This probability can be computed explicitly by generating trajectories from an isotropically distributed ensemble of configurations of particles at distance σ. To prevent needlessly long trajectories, Northrup et al. introduced an additional surface c at which they halt the trajectories. The computed association probability $\beta \neq \beta_\infty$ is now defined as the chance to associate rather than to escape to c. They then relate β to β_∞ using a branching method, adding up all probabilities of paths that leave c but return to σ rather than reach infinity[21], yielding a geometric series that can be written as

$$\beta_\infty = \frac{\beta}{1 - (1-\beta)\Omega}, \tag{4.29}$$

where $\Omega = k_D(\sigma)/k_D(c) = \sigma/c$. As the association probability β_∞ is related to the intrinsic rate constant through $\beta_\infty = k_a(\sigma)/(k_a(\sigma) + k_D(\sigma))$, it follows that

$$k_a(\sigma) = \frac{\beta k_D(\sigma)}{(1-\beta)(1-\Omega)}. \tag{4.30}$$

The link with the intrinsic rate constant expression Eq. 4.20 can be made by realizing that the association probability $\beta = 1 - P(c|\sigma)$, leading to

$$k_a(\sigma) = \frac{(1 - P(c|\sigma))\, k_D(\sigma)}{P(c|\sigma)(1-\Omega)}. \tag{4.31}$$

This expression is identical to Eq. 4.20, with $c = r_n$. Strikingly, while derived in a different way, the intrinsic rate expression based on the Northrup method yields the same correction factor as our analysis for finite interfaces.

4.7. ANISOTROPIC INTERACTIONS

The above derivations hold for particles interacting via an isotropic pair potential. However, many molecular systems, such as proteins and ligands, have anisotropic interactions that depend on the relative orientation of the particles. For such systems it is not immediately clear whether the expressions derived above are still valid. The point is that while the Boltzmann distribution of the particles' orientations is isotropic beyond the cutoff distance of the potential, the distribution in the ensemble of reactive trajectories, as harvested by TIS and FFS, is not: in these reactive trajectories, the particles tend to have their patches aligned. Naturally, one can still define and measure an effective association and dissociation rate. Yet, the simple expressions derived in the appendix are no longer valid.

Following Northrup et al. [8, 21], one can always express the effective association rate for anisotropic particles as

$$k_{\mathrm{on}} = \beta_\infty k_D(\sigma), \tag{4.32}$$

where the diffusion-limited rate constant is again $k_D(\sigma) = 4\pi D\sigma$, and β_∞ is given by Eq. 4.29. Now, we *define* the intrinsic rate $k_a(\sigma)$ via $\beta_\infty \equiv k_a(\sigma)/(k_a(\sigma) + k_D(\sigma))$, which yields

$$k_a(\sigma) = \frac{\beta k_D(\sigma)}{(1-\beta)(1-\Omega)}, \tag{4.33}$$

with Ω, as before, given by $\Omega = k_D(\sigma)/k_D(c)$. This yields an explicit expression for $k_a(\sigma)$ in terms of the probability β of binding rather than reaching the surface $r_n = c$, starting from an *isotropic* distribution at the surface σ; this is indeed the essence of the technique of Northrup et al. [21]. When the distribution as generated in the TIS/FFS simulation *is* isotropic at σ, then Eqs. 4.20 and 4.33 with $\beta = 1 - P(r_n|\sigma)$ are equivalent and both can be used. The problem arises when we want to connect Eq. 4.33 to the expression used in TIS/FFS to compute the dissociation rate in the case that the distribution at σ is not isotropic. The principal idea of the scheme presented in sections 4.4 and 4.5 is that $P(r_n|\sigma)$, as obtained in a TIS/FFS computation of the dissociation rate, is given by the analytical result $k_D(\sigma)/(k_a(\sigma) + k_D)$, for $r_n \to \infty$. We could thus use this analytical result to obtain the intrinsic rate $k_a(\sigma)$ in Eqs. 4.13 and 4.14; the expression that relates $P(r_n|\sigma)$ to $k_a(\sigma)$ when r_n is finite, is Eq. 4.20. However, in the case of anisotropic interaction potentials the distribution of reactive trajectories at the σ interface can also become anisotropic. In that case one can no longer identify $P(r_n|\sigma)$ as obtained in an TIS/FFS simulation with $1-\beta$ in Eq. 4.33, and Eq. 4.20 or Eq. 4.33 cannot be used to obtain $k_a(\sigma)$ from a TIS/FFS simulation. In summary, when the distribution at σ as obtained in the TIS/FFS simulation is isotropic, we expect both Eq. 4.33 and Eq. 4.20 to hold. However, if the distribution is not isotropic Eq. 4.20 ceases to be valid.

Nevertheless, even when the potential is anisotropic, Eq. 4.20 still provides a route towards computing $k_a(\sigma)$. Indeed, if at a certain interface σ' sufficiently far away from contact the distribution of trajectories has become uniform, the intrinsic rate for that surface is given by Eq. 4.20

$$k_a(\sigma') = \frac{(1 - P(r_n|\sigma'))k_D(\sigma')}{P(r_n|\sigma')(1-\Omega)}, \tag{4.34}$$

where $\Omega = k_D(\sigma')/k_D(r_n)$, and $r_n > \sigma'$. Since we know that the effective association rate is independent on the choice of the dividing surface, we can write

$$\frac{1}{k_{\text{on}}} = \frac{1}{k_a(\sigma)} + \frac{1}{k_D(\sigma)} = \frac{1}{k_a(\sigma')} + \frac{1}{k_D(\sigma')}, \tag{4.35}$$

even if the distribution at σ is anisotropic. Inserting Eq. 4.34 into this identity gives

$$\frac{1}{k_a(\sigma)} = \frac{P(r_n|\sigma')(1-\Omega)}{(1 - P(r_n|\sigma'))k_D(\sigma')} + \frac{1}{k_D(\sigma')} - \frac{1}{k_D(\sigma)}, \tag{4.36}$$

or, rearranging,

$$k_a(\sigma) = \frac{(1 - P(r_n|\sigma'))k_D(\sigma')k_D(\sigma)}{P(r_n|\sigma')(k_D(\sigma') - k_D(\sigma)\Omega) + k_D(\sigma) - k_D(\sigma')}. \tag{4.37}$$

This expression reduces reduces to Eq. 4.20 when $\sigma = \sigma'$, as expected. For $\sigma \neq \sigma'$, the expression holds even when the distribution at σ is anisotropic, provided that the distribution at σ' is isotropic. The value of $k_a(\sigma)$ thus obtained via a TIS/FFS computation of the dissociation pathway, yielding an anisotropic distribution at σ, is the same as that would have been obtained from a simulation of the association pathway via the Northrup scheme starting from a uniform distribution at σ.

Inserting Eq. 4.36 into Eq. 4.35 yields

$$k_{\text{on}} = \frac{(1 - P(r_n|\sigma'))k_D(\sigma')}{1 - \Omega P(r_n|\sigma')}, \tag{4.38}$$

which is indeed identical to Eq. 4.21 with $\sigma = \sigma'$. Since k_{on} is independent of the interface σ', the rate given by Eq. 4.38 as a function of σ' should reach a constant value. Any deviation from this limiting value is due to a loss of isotropy. Hence, this expression provides a criterion for testing the isotropy requirement.

4.7.1. ILLUSTRATIVE EXAMPLE

We consider a system comprising of two patchy particles interacting via a pair potential that consists of several contributions. The particles interact via a center-center potential $V_{cc}(r) = V_{\text{cc-rep}}(r) + V_{\text{cc-att}}(r)$ based on the distance r between the centers of mass, that is in turn built up from a repulsive and an attractive potential. Additionally, there is an attractive patch-patch pair interaction $V_{\text{pp}}(\delta)$ based on the distance δ between two points (patches) located on either particle's surface. Fig. 4.3 illustrates this setup. The total pair interaction potential is thus

$$V(r,\delta) = V_{\text{cc-rep}}(r) + V_{\text{cc-att}}(r) + V_{\text{pp}}(\delta). \tag{4.39}$$

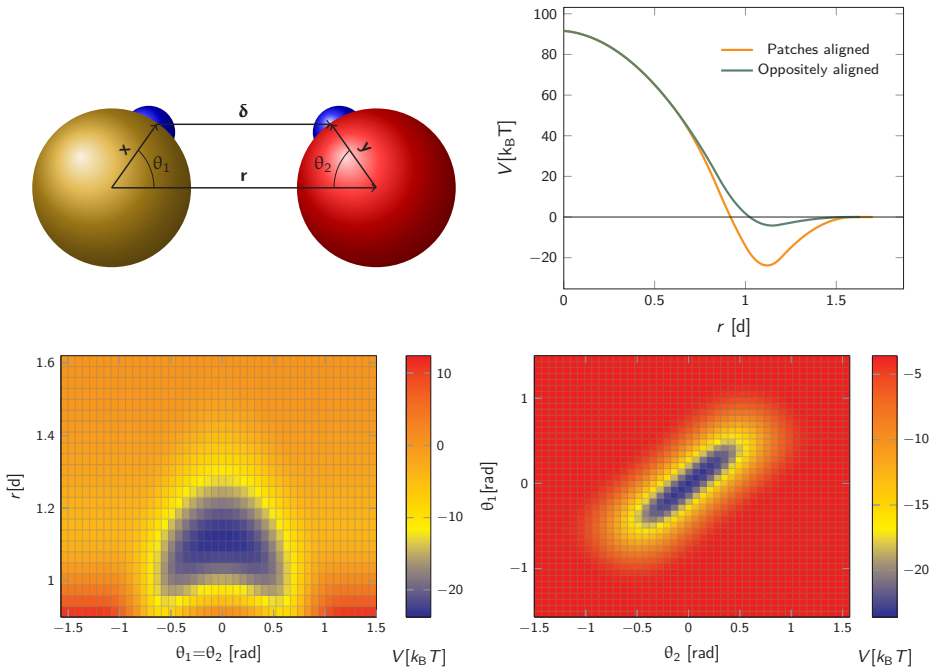


Figure 4.3: The anisotropic interaction potential. Top left: The particles interact via a combination of a center-center potential $V_{\mathrm{cc}}(r)$ and a patch-patch attraction $V_{\mathrm{pp}}(\delta)$ between the small patches on the surface of the particle. Top right: Total interaction potential $V_{\mathrm{cc}}(r) + V_{\mathrm{pp}}(r-d)$ for two particles with perfectly aligned complementary patches (orange solid curve), and oppositely aligned patches $V_{\mathrm{cc}}(r) + V_{\mathrm{pp}}(r+d)$ (dark green curve). The existence of patches introduces a strongly attractive bound state with the aligned particles in close contact. Bottom left: Heat map of the potential as a function of the distance between centres of mass r and the angle between the patch vector and inter-particle vector, θ_1 and θ_2 (see also top left) with $\theta_1 = \theta_2$ Bottom right: Heat map of potential as a function of θ_1, θ_2 given $r = 1.1d$. Note the relatively narrow range of orientations over which strong bonding occurs.

Note that δ implicitly depends on the location of the patch, and hence on the orientation of the particle. The potential $V_{\text{cc-rep}}(r)$, $V_{\text{cc-att}}(r)$ and $V_{\text{pp}}(\delta)$ have the simple quadratic form

$$V_i(x) = \begin{cases} \epsilon_i(1 - a_i\left(\frac{x}{d}\right)^2) & \text{if } x < x_i^\star, \\ \epsilon_i b_i(\frac{x_i^c}{d} - \frac{x}{d})^2 & \text{if } x_i^\star < x < x_i^c, \\ 0 & \text{otherwise.} \end{cases} \tag{4.40}$$

where the index i refers to one of the labels {cc-rep, cc-att, pp }, and d is the particle diameter and sets the length scale. The overall strength ϵ_i, 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^*. These potentials give us a firm control over the potential shape, and allow for easy integration with potentials that are short-ranged and highly orientation-specific. For our illustrative purposes, we take the following parameters: $\epsilon_{\text{cc-rep}} = 100k_BT$, $a_{\text{cc-rep}} = 1$ and $R^*_{\text{cc-rep}} = 0.85d$, implying $b_{\text{cc-rep}} = 2.6036$ and $R^c_{\text{cc-rep}} = 1.1764d$; $\epsilon_{\text{pp}} = 20k_BT$, $a_{\text{pp}} = 20$ and $r^*_{\text{pp}} = 0.1d$, implying $b_{\text{pp}} = 5$ and $r^c_{\text{pp}} = 0.5d$; and $\epsilon_{\text{cc-att}} = 10k_BT$, $a_{\text{cc-att}} = 1$ and $R^*_{\text{cc-att}} = 0.85d$, implying $b_{\text{cc-rep}} = 2.6036$ and $R^c_{\text{cc-rep}} = 1.1764d$.

As an illustration we plot in Fig. 4.3 the total inter-particle potential as a function of r when the two complementary patches are aligned to face each other, so that $\delta = r - d$. A narrow attractive well arises when the two particles are in close contact. In the same plot we also give the inter-particle potential as a function of r when the two complementary patches are oppositely aligned (with the patches facing in opposite directions) so that $\delta = r + d$. Here, there is only a very shallow attractive interaction at contact, caused by the isotropic center-center potential $V_{cc}(r)$. In Fig. 4.3 we also demonstrate the orientational dependence of the attractive potential, showing that the attractive interaction is highly sensitive to misalignment.

A Brownian Dynamics (BD) simulation employing an anisotropic potential requires translation as well rotational dynamics. We use a second order quasi-symplectic BD integrator which works particularly well for orientational dynamics [22], and can be straightforwardly combined with FFS. We performed an FFS simulation, with interfaces based on the binding energy when the particles are within the range of the potential [23], and beyond that with interfaces based on the center-to-center distance r: r ={1.5, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8 ,3.0, 3.5 , 5.5, 6.5, 7.5 } in units of d. From each interface, 10000 trajectories are started, and the conditional probability as in Eq. 4.8 is calculated.

To illustrate that Eq. 4.20 does not hold when the distribution at σ is not isotropic, we evaluated $k_a(\sigma)$ using this equation for a fixed cross section $\sigma = 1.5d$, see Fig. 4.4. The range of the anisotropic interaction potential depends on the orientation, but even when the patches are aligned (and the range is maximal), the potential approaches zero beyond $1.5d$ (see Fig. 4.3); this is indeed why we have chosen the cross-section to be $\sigma = 1.5d$. Fig. 4.4 shows $k_a(\sigma)$ as a function of the position of the last interface, r_n. It also shows the factors $1/(1-\Omega) = 1/(1-\sigma'/r_n)$ and $1 - P(r_n|\sigma')/P(r_n|\sigma')$, which both depend on r_n. However, while for the isotropic potential, $k_a(\sigma)$ is independent of r_n as long as σ (and r_n) are beyond the range of the interaction potential (see Fig. 4.1), here this is not the case, because the distribution at σ is not isotropic. This shows that Eq. 4.20 cannot

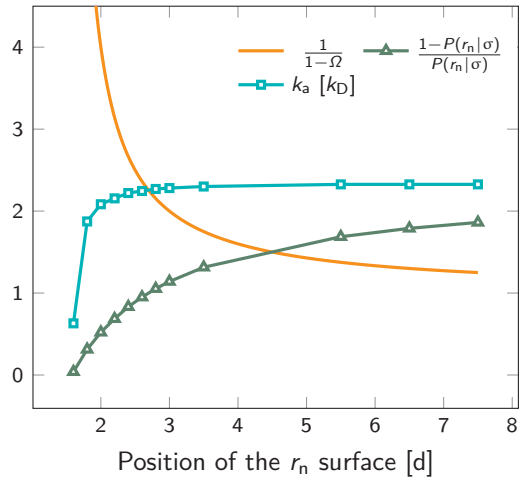


Figure 4.4: The intrinsic rate constant $k_a(\sigma)$ from Eq. 4.20 (in units of $k_D(\sigma)$) as a function of the position of the last interface, r_n, for the anisotropic interaction potential shown in Fig. 4.3. The cross section σ surface is kept constant at $1.5d$. The correction factor $1/(1-\Omega) = 1/(1-\sigma'/r_n)$ and the factor $(1-P(r_n|\sigma'))/P(r_n|\sigma')$ are also shown. In contrast to the behavior for the isotropic potential, Fig. 4.1, the value of $k_a(\sigma)$ depends on r_n, even when r_n is beyond the range of the potential $r > \sigma = 1.5d$. This is because for this anisotropic potential the distribution at σ is not uniform. As a result, Eq. 4.20 cannot be used.

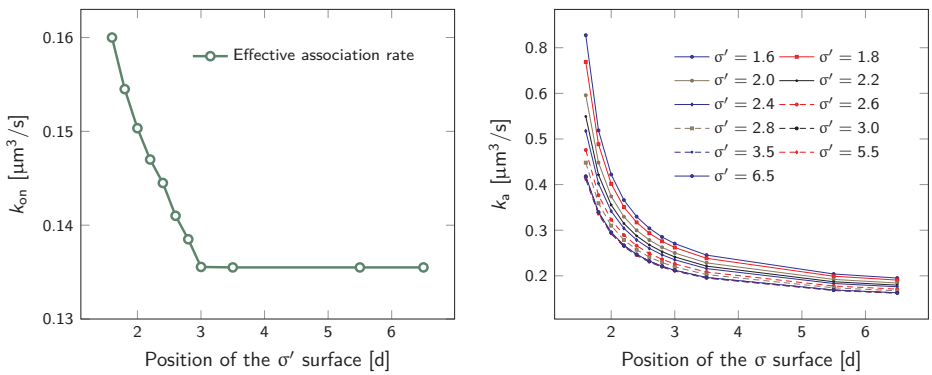


Figure 4.5: The effective and intrinsic association rates for the anisotropic potential of Fig. 4.3. Left: Effective association rate k_{on}, computed via Eq. 4.38, as a function of the position of the σ' surface, with the position of the last surface fixed at $r_n = 7.5d$. The value of k_{on} is constant for $\sigma' \geq 3d$, indicating that at and beyond this surface the particles are isotropically distributed. Right: The intrinsic rate constant $k_a(\sigma)$, computed via Eq. 5.12, as a function of the cross section σ for different positions of the σ' surface, with the position of the last interface fixed at $r_n = 7.5d$. The curves for $\sigma' = 3.0, 3.5, 5.5, 6.5$ overlap since the distribution of particles beyond this distance has become isotropic. For these values of σ', $k_a(\sigma)$ as a function of σ can be obtained.

be used for anisotropic interactions, if at σ the orientational distribution of the particles is still anisotopic.

However, plotting the effective rate k_{on} given by Eq. 4.38 as a function of σ' for $r_n = 7.5d$ in Fig. 4.5(left), shows that k_{on} becomes constant beyond $\sigma' = 3d$, indicating that this is the distance at which the particles become isotropically distributed. This observation allows us, via Eq. 5.12, to determine $k_a(\sigma)$, even when at σ the distribution is not isotropic. This is illustrated in Fig. 4.5(right), which shows $k_a(\sigma)$ as a function of σ from Eq. 5.12, for several values of σ', and for a fixed position of the last interface, $r_n = 7.5d$. The intrinsic rate shows qualitatively similar behavior as in Fig. 4.1, but for $\sigma' < 3d$, the value of $k_a(\sigma)$ does depend on σ', because the distribution at σ' is not isotropic yet. However, for $\sigma' > 3d$, $k_a(\sigma)$ becomes independent of σ', and the intrinsic rate constant k_a can thus be obtained for all values of σ that are beyond the range of the interaction potential.

4.8. CONCLUSION

In this work we derived explicit microscopic expressions for the intrinsic rate constants for diffusion influenced reactions. Remarkably all intrinsic and effective rate constants, as well as the equilibrium constant can be computed from a single TIS or FFS simulation of the dissociation process. To the best of our knowledge this is a new result, and has not been reported in the literature before. We illustrated that this approach works for generic isotropic potentials and even for anisotropic potentials when the reference interface is sufficiently far from contact such that the orientational distributions are isotropic. This later condition led to a criterion for testing this isotropic behavior.

The results obtained in this paper are very general and can be used to compute accurate rate constants in complex systems using rare events techniques such as Forward Flux Sampling or Transition Interface Sampling. In future work we will also study how the magnitude of the intrinsic rate constant compares to that of the diffusion-limited rate constant, and how this depends on the binding affinity. This question is, for instance, important for understanding how tightly diffusion puts a fundamental upper bound on the precision of chemical sensing [24, 25, 11, 26].

ACKNOWLEDGEMENTS

We thank Tom Ouldridge for stimulating discussions and help with implementing the LD integrator. 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.

APPENDIX A: THE RATE CONSTANTS FOR TWO PARTICLES THAT INTERACT VIA AN ISOTROPIC INTERACTION POTENTIAL

It will be instructive to first revisit the derivation of the rate constants. Following Agmon and Szabo [1], we consider a single static receptor at the origin and a single ligand molecule that moves with diffusion constant D. The probability that the ligand molecule is at distance r at time t given that it was initially at a distance r_0 is given by the Green's function $p(r,t|r_0)$. The evolution of the Green's function is given by the diffusion equation

$$\frac{\partial p(r,t|r_0)}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r} D r^2 e^{-\beta U(r)} \frac{\partial}{\partial r} e^{\beta U(r)} p(r,t|r_0), \tag{4.41}$$

where β is the inverse temperature and $U(r)$ is the interaction potential. The effective association and dissociation rate constants are obtained by solving Eq. 4.41 with different boundary conditions. We start with the association reaction.

ASSOCIATION REACTION

To obtain the effective association rate constant k_{on}, we solve Eq. 4.41 subject to the boundary condition

$$4\pi\sigma^2 D \left.\frac{\partial p(r,t|r_0)}{\partial r}\right|_{r=\sigma} = k_a p(\sigma,t|r_0). \tag{4.42}$$

Here k_a is the intrinsic rate constant, which determines the rate at which receptor and ligand associate given they are at the contact distance σ. If k_a is finite, then the boundary condition is called a radiation boundary condition, while if $k_a \to \infty$, the boundary condition is an absorbing condition. The latter can be used to obtain the rate constant of diffusion-limited reactions, where receptor and ligand associate upon the first collision.

The survival probability $S_\alpha(t|r_0)$ is the probability that a particle, which starts at a position r_0, has not yet reacted at a later time t. It is given by

$$S_\alpha(t|r_0) = 4\pi \int_\sigma^\infty dr\, r^2 p(r,t|r_0). \tag{4.43}$$

The subscript α is either "rad" or "abs", corresponding to k_a being finite or infinite, respectively. The propensity function $R_\alpha(t|r_0)$ is the probability that a ligand particle, which starts at $r = r_0$, reacts for the first time at a later time t:

$$R_\alpha(t|r_0) = -\frac{\partial S_\alpha(t|r_0)}{\partial t}. \tag{4.44}$$

The time-dependent rate constant $k_\alpha(t)$ is

$$k_\alpha(t) = 4\pi \int_\sigma^\infty dr_0\, r_0^2 R_\alpha(t|r_0) p_{\text{eq}}(r_0). \tag{4.45}$$

The distribution $p_{\text{eq}}(r_0)$ is the equilibrium radial distribution function, $p_{\text{eq}}(r) = e^{-\beta U(r)}$. If ligand and receptor only interact at contact, then $U(r) = 0$ for $r \geq \sigma$ and $p_{\text{eq}} = 1$, meaning that the equilibrium distribution corresponds to a spatially uniform distribution.

The time-dependent rate constant $k_\alpha(t)$ divided by the volume V is the probability per unit amount of time that receptor and ligand associate for the first time at a later time t, averaged over all initial positions r_0 drawn from the equilibrium distribution $p_{\rm eq}(r_0)$.

The expressions Eq. 4.43 - Eq. 4.45 hold for both radiating and absorbing boundary conditions, corresponding to k_a being finite and infinite, respectively. When k_a is finite, $R_{\rm rad}(t|r_0)$ is also given by

$$R_{\rm rad}(t|r_0) = k_a p(\sigma, t|r_0) \tag{4.46}$$

and the time-dependent rate constant $k_{\rm rad}(t)$ is then also given by

$$k_{\rm rad}(t) = 4\pi k_a \int_\sigma^\infty dr_0 r_0^2 p(\sigma, t|r_0) p_{\rm eq}(r_0) \tag{4.47}$$

To relate $k_{\rm rad}(t)$ to $k_{\rm abs}(t)$ in what follows below it will be useful to exploit the detailed-balance condition

$$p_{\rm eq}(r_0) p(r, t|r_0) = p_{\rm eq}(r) p(r_0, t|r). \tag{4.48}$$

We can integrate this equation over r_0 to find

$$4\pi \int dr_0 r_0^2 p(r, t|r_0) p_{\rm eq}(r_0) = p_{\rm eq}(r) S_\alpha(t|r). \tag{4.49}$$

Combining this equation with Eq. 4.47 we find that

$$k_{\rm rad}(t) = p_{\rm eq}(\sigma) k_a S_{\rm rad}(t|\sigma). \tag{4.50}$$

The time-dependent rate constant $k_{\rm rad}(t)$ can be related to the time-dependent rate constant $k_{\rm abs}(t)$ via

$$k_{\rm rad}(t) = \int_0^t dt' R_{\rm rad}(t - t'|\sigma) k_{\rm abs}(t'). \tag{4.51}$$

This can be understood by noting that $k_{\rm abs}(t')/V$ is the probability per unit amount of time that receptor and ligand come in contact for the first time at time t', while $R_{\rm rad}(t - t'|\sigma)$ is the probability that receptor and ligand which start at contact $r = \sigma$ at time t' associate a time $t - t'$ later. In Laplace space, the above expression reads

$$\hat{k}_{\rm rad}(s) = \hat{R}_{\rm rad}(s|\sigma) \hat{k}_{\rm abs}(s). \tag{4.52}$$

Since $R_{\rm rad}(t|\sigma) = -\partial S_{\rm rad}(t|\sigma)/\partial t$, $\hat{R}_{\rm rad}(s|\sigma)$ is also given by

$$\hat{R}_{\rm rad}(s|\sigma) = 1 - s\hat{S}_{\rm rad}(s|\sigma). \tag{4.53}$$

Combining the Laplace transform of Eq. 4.50 with Eq. 4.52 and Eq. 4.53 yields

$$\hat{k}_{\rm rad}(s) = \frac{k_a p_{\rm eq}(\sigma) \hat{k}_{\rm abs}(s)}{k_a p_{\rm eq}(\sigma) + s\hat{k}_{\rm abs}(s)}. \tag{4.54}$$

The effective association rate $k_{\rm on}$ is given by the long-time limit of $k_{\rm rad}(t)$. Using Eq. 4.54 we thus find:

$$k_{\rm on} = \lim_{t\to\infty} k_{\rm rad}(t) = \lim_{s\to 0} s\hat{k}_{\rm rad}(s) = \frac{k_a p_{\rm eq}(\sigma) k_D}{k_a p_{\rm eq}(\sigma) + k_D}. \tag{4.55}$$

Here, $k_D = \lim_{s\to 0} s\hat{k}_{\rm abs}(s)$ is the diffusion-limited association rate for two particles interacting via an interaction potential $U(r)$. By combining Eq. 4.50 with Eqs. 4.52 and 4.53, it also follows that the probability that a particle at contact σ does not bind but escapes, is

$$S_{\rm rad}(\infty|\sigma) = \lim_{s\to 0} s\hat{S}(s|\sigma) = \frac{k_D}{k_a p_{\rm eq}(\sigma) + k_D}. \tag{4.56}$$

Hence, the effective association rate constant, Eq. 4.55, can be interpreted as being given by the rate constant k_D of arriving at the surface σ, followed by the probability $1-S_{\rm rad}(\infty|\sigma) = k_a p_{\rm eq}(\sigma)/(k_a p_{\rm eq}(\sigma) + k_D)$ that the arrival leads to binding.

The results derived above hold for arbitrary $p_{\rm eq}(r) = e^{-\beta U(r)}$. We now consider the case that $U(r) = 0$ for $r \geq \sigma$. The time-dependent rate constant $k_{\rm abs}(t)$ is then [27]

$$k_{\rm abs}(t) = 4\pi\sigma D\left(1 + \sigma/\sqrt{\pi D t}\right), \tag{4.57}$$

which in the Laplace domain becomes

$$s\hat{k}_{\rm abs}(s) = k_D\left(1 + \tau(s)\right), \tag{4.58}$$

where $\tau(s) \equiv \sigma\sqrt{s/D} = \sqrt{s\tau_m}$ with the molecular time scale $\tau_m = \sigma^2/D$ and $k_D \equiv k_{\rm abs}(t\to\infty) = 4\pi\sigma D$ is the diffusion-limited rate constant for two particles that do not interact except at contact. Substituting this in Eq. 4.54 with $U(\sigma) = 0$ gives

$$\hat{k}_{\rm rad}(s) = \frac{k_a k_D}{s}\frac{1+\tau(s)}{k_a + k_D(1+\tau(s))}. \tag{4.59}$$

This expression yields for the effective association rate $k_{\rm on} = \lim_{s\to 0} s\hat{k}_{\rm rad}(s)$:

$$k_{\rm on} = \frac{k_a k_D}{k_a + k_D}. \tag{4.60}$$

We note that this expression also follows from the much simpler steady state approximation of the macroscopic rate equations for association, in which the surface σ is viewed as an intermediate state. In this approximation k_a is taken as the rate from the intermediate σ surface to the associated/bound state and k_D is the diffusion limited rate to reach the σ surface from the unbound state. However, the above derivation is more general, and does not require the (rather strong) approximations that are made in the simple approach.

THE DISSOCIATION RATE CONSTANT

Following Amgon and Szabo [1], we will derive the effective dissociation rate constant k_{off} from $S_{\text{rev}}(t|*)$:

$$\frac{1}{k_{\text{off}}} = \tau_{\text{off}} = \int_0^\infty dt\,[1 - S_{\text{rev}}(t|*)]\,. \tag{4.61}$$

Here $S_{\text{rev}}(t|*)$ is the probability that the ligand, which is bound initially, is free at a later time t. The subscript "rev" indicates that during the time t the ligand may bind and unbind many times.

To obtain $S_{\text{rev}}(t|*)$, we first consider $S_{\text{rev}}(t|r_0)$ and the following boundary condition for Eq. 4.41 [1]:

$$\begin{aligned} 4\pi\sigma^2 D\,\frac{\partial p_{\text{rev}}(r,t|r_0)}{\partial r}\bigg|_{r=\sigma} &= k_a p_{\text{rev}}(\sigma,t|r_0) - k_d\,[1 - S_{\text{rev}}(t|r_0)] \\ &= R_{\text{rev}}(t|r_0). \end{aligned} \tag{4.62}$$

Using $\hat{R}_{\text{rev}}(s|r_0) = 1 - s\hat{S}_{\text{rev}}(s|r_0)$ we can rewrite the above boundary condition as

$$\hat{R}_{\text{rev}}(s|r_0) = \frac{sk_a}{s+k_d}\,\hat{p}_{\text{rev}}(\sigma,s|r_0). \tag{4.63}$$

The rate is, analogous to Eq. 4.45,

$$k_{\text{rev}}(t) = 4\pi \int dr_0\, r_0^2 R_{\text{rev}}(t|r_0) p_{\text{eq}}(r_0). \tag{4.64}$$

Using the detailed-balance relation Eq. 4.49, this equation can be combined with Eq. 4.63, to give

$$\hat{k}_{\text{rev}}(s) = \frac{sk_a p_{\text{eq}}(\sigma)}{s+k_d}\,\hat{S}_{\text{rev}}(s|\sigma). \tag{4.65}$$

Combining this equation with $\hat{k}_{\text{rev}}(s) = \hat{R}_{\text{rev}}(s|\sigma)\hat{k}_{\text{abs}}(s)$ and $\hat{R}_{\text{rev}}(s|\sigma) = 1 - s\hat{S}_{\text{rev}}(s|\sigma)$, we can derive that

$$\hat{k}_{\text{rev}}(s) = \frac{k_a p_{\text{eq}}(\sigma)\hat{k}_{\text{abs}}(s)}{k_a p_{\text{eq}}(\sigma) + (s+k_d)\hat{k}_{\text{abs}}(s)}. \tag{4.66}$$

We now consider $S_{\text{rev}}(t|*)$. Since $S_{\text{rev}}(0|*) = 0$, $\hat{R}(s|*) = -s\hat{S}_{\text{rev}}(s|*)$. The boundary condition, Eq. 4.62, then becomes

$$\hat{R}_{\text{rev}}(s|*) = \frac{sk_a\hat{p}_{\text{rev}}(\sigma,s|*) - k_d}{s+k_d}. \tag{4.67}$$

Exploiting the Laplace transform of the detailed balance condition $k_a p_{\text{rev}}(\sigma,t|*) = k_d\, p_{\text{rev}}(*,t|\sigma)$ and the definition $p_{\text{rev}}(*,t|\sigma) \equiv 1 - S_{\text{rev}}(t|\sigma)$, the above equation yields

$$\hat{S}_{\text{rev}}(s|*) = \frac{k_d}{s+k_d}\,\hat{S}_{\text{rev}}(s|\sigma). \tag{4.68}$$

In the time domain, this gives

$$S_{\text{rev}}(t|*) = k_d \int_0^t \exp(-k_d t') S_{\text{rev}}(t-t'|\sigma) dt'. \tag{4.69}$$

which can be understood by noting that $k_d \exp(-k_d t')$ is the probability per unit amount of time that the bound ligand dissociates at time t' and $S_{\text{rev}}(t-t'|\sigma)$ is the probability that the dissociated particle, which is now at contact σ, is unbound at a later time $t-t'$ (but it could have associated and dissociated in between multiple times).

Combining Eq. 4.68 with Eq. 4.65 gives

$$\hat{k}_{\text{rev}}(s) = sK_{\text{eq}}\hat{S}_{\text{rev}}(s|*). \tag{4.70}$$

where $K_{\text{eq}} \equiv k_a p_{\text{eq}}(\sigma)/k_d$. Combining this result with Eq. 4.66, we find

$$s\hat{S}_{\text{rev}}(s|*) = \frac{k_d k_{\text{abs}}(s)}{k_a p_{\text{eq}}(\sigma) + (s+k_d)k_{\text{abs}}(s)}. \tag{4.71}$$

The off rate is then

$$\frac{1}{k_{\text{off}}} = \tau_{\text{off}} = \int_0^\infty dt\,[1-S(t|*)] = \lim_{s\to 0}\left[1/s - \hat{S}_{\text{rev}}(s|*)\right] = \frac{1}{k_d} + \frac{K_{\text{eq}}}{k_D}, \tag{4.72}$$

where, as before, k_D is the long-time limit of $k_{\text{abs}}(t)$: $k_D = \lim_{s\to 0} sk_{\text{abs}}(s)$. This result can be rewritten as

$$k_{\text{off}} = \frac{k_d k_D}{k_a p_{\text{eq}}(\sigma) + k_D}, \tag{4.73}$$

which, using Eq. 4.56, is also given by

$$k_{\text{off}} = k_d S_{\text{rad}}(\infty|\sigma). \tag{4.74}$$

Indeed, the effective off rate is the intrinsic dissociation rate k_d times the probability $S_{\text{rad}}(\infty|\sigma) = k_D/(k_a p_{\text{eq}}(\sigma) + k_D)$ that the particle subsequently escapes. As written, the above results for k_{off} hold for any $U(r)$. When $U(r) = 0$ for $r \geq \sigma$, $p_{\text{eq}}(\sigma) = 1$ and $k_D = 4\pi\sigma D$.

Finally, we note that, from Eqs. 4.55 and 4.73, it is clear that the equilibrium constant is also given by

$$K_{\text{eq}} = \frac{k_a p_{\text{eq}}(\sigma)}{k_d} = \frac{k_{\text{on}}}{k_{\text{off}}}. \tag{4.75}$$

BIBLIOGRAPHY

[1] Agmon N, Szabo A (1990) Theory of reversible diffusion-influenced reactions. *The Journal of Chemical Physics* 92:5270.

[2] van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *The Journal of chemical physics* 123:234910.

[3] Morelli MJ, ten Wolde PR (2008) Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network. *The Journal of Chemical Physics* 129:054112.

[4] Takahashi K, Tănase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America* 107:2473–8.

[5] Lipková J, Zygalakis KC, Chapman SJ, Erban R (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. *J. Appl. Math.* 71:714–730.

[6] Johnson ME, Hummer G (2014) Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. *Phys. Rev. X* 4:031037.

[7] Klein HCR, Schwarz US (2014) Studying protein assembly with reversible Brownian dynamics of patchy particles. *The Journal of Chemical Physics* 140:184112.

[8] Northrup SH, Erickson HP (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proceedings of the National Academy of Sciences* 89:3338–3342.

[9] Zhou HX (1998) Comparison of three Brownian-dynamics algorithms for calculating rate constants of diffusion-influenced reactions. *The Journal of Chemical Physics* 108:8139.

[10] van Zon JS, Morelli MJ, Tanase-Nicola S, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expresssion. *Biophysical Journal* 91:4350.

[11] Kaizu K, et al. (2014) The berg-purcell limit revisited. *Biophysical journal* 106:976–985.

[12] Gopich IV, Szabo A (2013) Diffusion modifies the connectivity of kinetic schemes for multisite binding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 110:19784–19789.

[13] Mugler A, ten Wolde PR (2013) The Macroscopic Effects of Microscopic Heterogeneity in Cell Signaling. *Advances in Chemical Physics* 153:373–396.

[14] Mugler A, Tostevin F, ten Wolde PR (2013) Spatial partitioning improves the reliability of biochemical signaling. *Proceedings of the National Academy of Sciences of the United States of America* 110:5927–5932.

[15] van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. *The Journal of Chemical Physics* 118:7762.

[16] van Erp TS, Bolhuis PG (2005) Elaborating transition interface sampling methods. *J. Comp. Phys.* 205:157–181.

[17] van Erp TS (2012) Dynamical rare event simulation techniques for equilibrium and nonequilibrium systems. *Adv. Chem. Phys.* 151:27–60.

[18] Allen RJ, Warren PB, ten Wolde PR (2005) Sampling Rare Switching Events in Biochemical Networks. *Physical Review Letters* 94:018104.

[19] Allen RJ, Valeriani C, Rein ten Wolde P (2009) Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter* 21:463102.

[20] Von Smoluchowski M (1917) *Z. Phys. Chem.* 92:129.

[21] Northrup SH, Allison SA, McCammon JA (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80:1517.

[22] Vijaykumar A, Bolhuis PG, ten Wolde PR (2015) Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations. *The Journal of Chemical Physics* 143:214102.

[23] Vijaykumar A, Ouldridge TE, ten Wolde PR, Bolhuis PG (2016) Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulation for patchy particles. *To be submitted.*

[24] Berg HC, Purcell EM (1977) Physics of chemoreception. *Biophysical Journal* 20:193.

[25] Bialek W, Setayeshgar S (2005) Physical limits to biochemical signaling. *Proceedings of the National Academy of Sciences USA* 102:10040–10045.

[26] ten Wolde PR, Becker NB, Ouldridge TE, Mugler A (2016) Fundamental Limits to Cellular SensingPieter Rein ten Wolde, Nils B. Becker, Thomas E. Ouldridge & Andrew Mugler. *Journal of Statistical Physics* 162:1395–1424.

[27] Rice SA (1985) *Diffusion-limited reactions* eds Bamford C, Tipper C, Compton R (Elsevier, Amsterdam).

5

THE MAGNITUDE OF THE INTRINSIC RATE CONSTANT: HOW DEEP CAN ASSOCIATION REACTIONS BE IN THE DIFFUSION LIMITED REGIME?

Intrinsic and effective rate constants have an important role in the theory of diffusion-limited reactions. In a previous paper, we provide detailed microscopic expressions for these intrinsic rates [1], which are usually considered as abstract quantities and assumed to be implicitly known. Using these microscopic expressions, we investigate how the rate of association depends on the strength and the range of the isotropic potential, and the strength of the non-specific attraction in case of the anisotropic potential. In addition, we determine the location of the interface where these expressions become valid for anisotropic potentials. In particular, by investigating the particles' orientational distributions we verify whether the interface at which these distributions become isotropic agrees with the interface predicted by the effective association rate constant. Finally, we discuss how large the intrinsic association rate can become, and what the consequences are for the existence of the diffusion limited regime.

5.1. INTRODUCTION

Association and dissociation of pairs of particles plays a central role, not only in cellular processes such as the binding of a ligand to a receptor, an enzyme to its substrate, or of a protein to DNA in gene regulation, but also in the self-assembly of colloids, in microemulsion formation or in the phase behavior of polymer solutions. During association particles come into contact via diffusion, and bind with a rate depending on the intrinsic association rate constant. When dissociating, a bound particle pair separates with an intrinsic dissociation rate, and after which the particles diffuse away from each other. Theories of diffusion-influenced reactions express the experimentally important *effective* rate constants in terms of diffusion constants, cross section, interaction potential of the particles, and the intrinsic association and dissociation rate constants [2]. The latter are often assumed as *a priori* given.

While simulation techniques exist that can compute association and dissociation rate constants for arbitrarily complex interaction potentials [3, 4], these are typically *effective* rate constants, resulting from both the diffusion process, and the binding rate upon contact due to the interaction potential. Since the effective rate is also what is often measured in experiments, few studies have focused on the dependence of intrinsic rate constants on the interaction potential, the cross section, and the particles' diffusion constants.

Knowledge of the effective association and dissociation rates is sufficient for describing the dynamics of dilute systems when the association and dissociation can be reduced to a two-body problem. Yet, it should be realized that, in general, association-dissociation reactions present a complicated non-Markovian many-body problem due to non-trivial spatio-temporal correlations between the reactions, such as rebinding events. In dilute systems it is often possible to integrate out the dynamics at the molecular scale, and describe association-dissociation as a Markovian process with effective rates describing the long time dynamics [5, 6]. However, the evidence is accumulating that even in dilute systems molecular-scale spatio-temporal correlations can dramatically influence the system's macroscopic scale behavior. For example, it has been predicted [7, 8] and shown by experiments [9] that in cellular systems which rely on multi-site protein modification, enzyme-substrate rebindings at the micrscopic scale can dramatically change the macroscopic behavior at the cellular scale. Also in chemical sensing, either by living cells or man-made sensors [10, 11, 12, 13, 14], the accuracy of sensing is affected by the microscopic dynamics of the ligand binding to and hopping between multiple receptors. In these cases, the effective rate constants are not sufficient for describing the dynamics: instead, this requires knowledge of diffusion constants, the cross sections, and the intrinsic rate constants. Also for the modeling of reactions in spatially heterogeneous cellular systems as well as in confined geometries or reduced dimensions, knowledge of the intrinsic rate constants is required. In addition, in recent years much effort has been devoted to understanding how the accuracy of sensing is set by the diffusion-limited arrival of the ligand at the receptor [10, 11, 12, 15, 13, 6, 16, 17, 14]. However, the limit derived is only tight if the ligand-receptor association reactions are indeed diffusion limited, meaning that the intrinsic rate constant is much larger than the diffusion-limited rate constant. Last but not least, knowledge of intrinsic rates is essential for simulations techniques that model (bio)chemical networks of chemical reactions

via reaction-diffusion, in which particles move by diffusion, and react at contact with a *given* intrinsic rate constant [18, 19, 7, 20, 21, 22, 23, 1, 24].

In Ref. [1] we derived microscopic expressions for these intrinsic rates that can be evaluated by focusing on a single dissociation reaction using a rare-event sampling technique. We demonstrated that these expressions not only hold for generic isotropic potentials, but also for anisotropic potentials, provided the cross section where the intrinsic rates are calculated is sufficiently far from contact that the particles' orientational distributions are isotropic.

In this work we use the technique introduced in Ref. [1] to study the intrinsic and effective rate constants as a function of the parameters of both isotropic and anisotropic (patchy) interaction potentials. For the isotropic potential we use the potential introduced in Ref. [25], which can accurately describe the phase behavior of small globular proteins such as lysozyme. For this potential, we investigate how the intrinsic and effective rate constant vary with the strength and the range of the interaction potential. For the anisotropic potential, we use the potential introduced in Ref. [1], which provides an effective coarse-grained model for proteins with patchy binding sites. The particles experience a repulsive force based on the distance between their centres of mass and a strong specific attractive force if their patches are aligned. In addition, particles also experience a weaker, isotropic attractive force, which models the non-specific binding in proteins. Besides the bound and unbound state, this isotropic non-specific attraction facilitates an intermediate weakly bound state, in which the particles are not bound to a specific patch but are likely to remain close to each other, increasing the chances that they rebind to a patch. We compute for this system how the intrinsic and effective rate constants vary with the non-specific attraction strength. Since the expressions for the rates derived in Ref. [1] are valid for the anisotropic potential only at cross sections sufficiently far from contact, we also determine the interface where the orientational distributions become isotropic as a function of the non-specific attraction strength.

The article is organized as follows: In the methods section, we present a brief derivation of the expressions for the rate constants using rare-event sampling techniques, followed by a short description of the particle model and the interaction potential and the methods that are used to perform the simulations. In the results section we evaluate these rates for the isotropic potential, varying its range and strength. Not surprisingly, the intrinsic and effective association rates increase with the interaction strength, while the corresponding dissociation rates strongly decrease. More interestingly, the intrinsic and effective association rates fall as the interaction range decreases, while the corresponding dissociation rate constants increase. We then determine the cross section beyond which these expressions hold for anisotropic potentials and calculate the anisotropic rates as a function of the non-specific attraction strength. The results show that non-specific binding can significantly speed up association and slow down dissociation. We end with a concluding section in which we discuss the question how deep the association process can be in the diffusion limited regime. For the potentials studied here the intrinsic association rate is about a factor 10 higher than the effective rate constant, which means that the overall association speed is more limited by diffusion than by the binding rate at contact, but certainly is not entirely dominated by it.

5.2. METHODS

5.2.1. INTRINSIC AND EFFECTIVE RATE CONSTANTS

In this section we briefly recapitulate the derivation that led up to the expressions for the intrinsic rate constants [1]. The central quantities in a reaction-diffusion system, as shown in Fig. 5.1, are: (i) the relative diffusion constant, D, which is the diffusion constant associated with the interparticle vector of the two associating species A and B: $D = D_A + D_B$, where D_A and D_B are the dissociation constants of the respective particles [26, 7]; (ii) the cross section σ; this cross section is the "contact distance" at which the particles can associate with each other; (iii) the intrinsic association rate, k_a, at which the particles can go from the contact state to the bound state; (iv) the intrinsic dissociation rate, k_d at which the particles can unbind from the bound state to the contact state; (v) the interaction potential $U(r)$.

When the intereaction potential $U(r)$ is zero beyond the constant distance σ, the rate at which the particles come into contact, starting from an equilibrium distribution, is given by the diffusion-limited rate constant, $k_D = 4\pi\sigma D$. In this case, the long-time limit of the time-dependent association rate is given by the effective association rate [2]:

$$k_{on} = k_D(\sigma).\frac{k_a(\sigma)}{k_a(\sigma) + k_D(\sigma)}. \tag{5.1}$$

This expression shows that that the effective association rate k_{on} can be interpreted as the rate at which particles come into contact, $k_D(\sigma)$, times the probability that given they are in contact, they also react, $k_a(\sigma)/(k_a(\sigma) + k_D(\sigma))$. We also emphasize that while k_{on} is independent of the choice of the contact distance σ, k_a and k_D do depend on this choice.

In the same limit $U(r = 0)$ for $r > \sigma$, the effective dissociation rate is given by [2]:

$$k_{off} = k_d(\sigma).\frac{k_D(\sigma)}{k_a(\sigma) + k_D(\sigma)}. \tag{5.2}$$

The effective dissociation rate k_{off} is thus the rate at which the bound particles dissociate, $k_d(\sigma)$, times the probability that they subsequently escape to infinity. While the effective dissociation rate k_{off} is independent of the choice of σ, the intrinsic dissociation rate $k_d(\sigma)$ does depend on it.

Dissociation is a rare event when particles are bound to patches, owing to strong specific attractive forces. Rare event sampling techniques such as Transition Interface Sampling [27] (TIS) or Forward Flux Sampling [28, 29] (FFS) allow evaluation of the corresponding low rate constants. The effective dissociation reaction k_{off} can be expressed as $k_{off} = \Phi_0 P(\infty|r_0)$, where Φ_0 is the flux of trajectories coming from the bound state, and crossing the r_0-interface for the first time, and $P(\infty|r_0)$ is the conditional probability for trajectories starting at the r_0-interface to reach infinity (unbound state), instead of returning back to the bound state. This probability of escaping to the unbound state from the r_0-interface can in turn be written as the probability of escaping from the bound state to the σ-interface first, multiplied by the probability of subsequently escaping to the unbound state from the σ-interface. It follows that

$$k_{off} = \Phi_0 P(\sigma|r_0) P(\infty|\sigma). \tag{5.3}$$

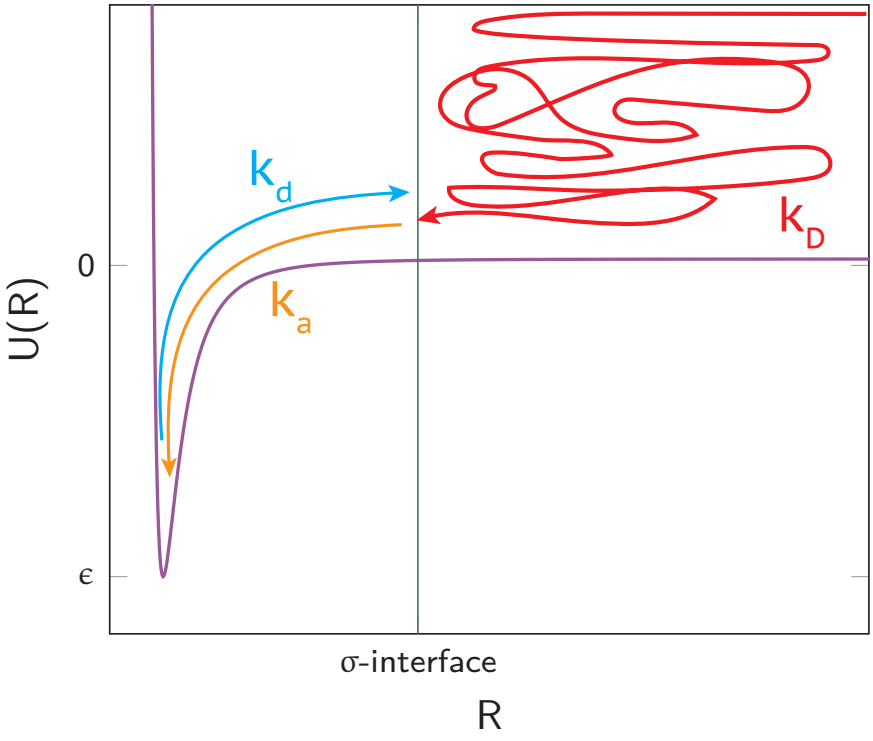


Figure 5.1: Central quantities in a reaction-diffusion system. Particles from an equilibrium distribution arrive at the contact interface, σ, with a diffusion limited rate, k_D. In the case of particles freely diffusing with a diffusion constant D, the analytical expression for the diffusion limited rate, $k_D = 4\pi\sigma D$. From contact the particles can go to the bound state with an intrinsic association rate, k_a and once bound the particles can unbind with an intrinsic dissociation rate, k_d. R is the inter-particle distance and $U(R)$ is the interaction potential.

Note that this expression is exact [27], and applies to any interaction potential. More specifically, it does not require that $U(r)$ is zero beyond the cut-off distance σ. The first two terms of Eq. 6.4 can be recognized as the definition of the intrinsic dissociation constant

$$k_d = \Phi_0 P(\sigma|r_0), \tag{5.4}$$

which makes it possible to write the effective dissociation constant as

$$k_{off} = k_d P(\infty|\sigma). \tag{5.5}$$

This expression is also exact, applying even when $U(r)$ is non-zero beyond σ. However, if the contact distance σ is chosen beyond the cut-off distance where the potential becomes zero, then we can combine Eq. 5.5 with Eq. 5.2 to express the escape probability in terms of the intrinsic association rate and the diffusion-limited rate constant:

$$P(\infty|\sigma) = \frac{k_D(\sigma)}{k_a(\sigma) + k_D(\sigma)}. \tag{5.6}$$

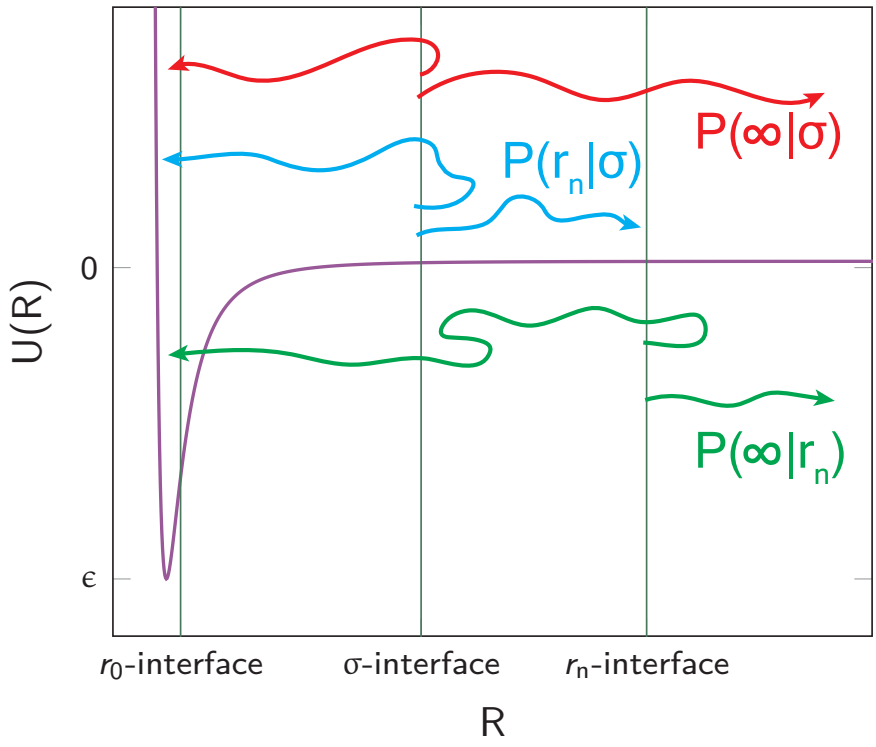


Figure 5.2: The main interfaces used in the FFS/TIS simulation, to calculate the rate constants. Particles within the r_0-interface are in a bound state and beyond the r_n-interface are in the unbound state. The σ-interface which is in between the r_0 and the r_n interfaces, is the cross-section where the particles are at contact. The conditional probabilities in the case of FFS/TIS is the probability of going from one interface to the next, as opposed to going back to r_0. R is the inter-particle distance and $U(R)$ is the interaction potential.

Rearranging leads to an expression for the intrinsic association rate

$$k_a(\sigma) = k_D(\sigma)\frac{1 - P(\infty|\sigma)}{P(\infty|\sigma)}. \tag{5.7}$$

$P(\infty|\sigma)$ could in principle be obtained by performing a single TIS/FFS simulation for the dissociation reaction. Hence, by putting σ beyond the cut-off distance of the potential, and by exploiting the analytical expressions for the diffusion-limited rate, $k_D(\sigma) = 4\pi\sigma$, and the escape probability, $P(\infty|\sigma) = k_D(\sigma)/(k_a(\sigma) + k_D(\sigma)$, we can obtain from one TIS/FFS simulation of a dissociation reaction, not only obtain the intrinsic rate $k_d(\sigma)$, via Eq. 6.3, and the effective dissociation rate k_{off}, via Eq. 5.5, but also the intrinsic association rate $k_a(\sigma)$, via Eq. 5.7, and the effective association rate k_{on}, via Eq. 5.1. Indeed, in all analyses performed below, we always put the cross section beyond the cutoff of the potential.

$P(\infty|\sigma)$ cannot be comnputed directly within a FFS/TIS simulation, because the last

interface can not be put at infinity in practice. In the simulations, we thus put the last interface at a finite distance r_n, and compute $P(r_n|\sigma)$ instead. However, we then need a correction, because the particles can rebind; in fact, even when r_n is chosen to be large, we still need this correction, because the rebinding probability decays, as we will see, only very slowy with r_n, namely as $1/r_n$. To relate $P(\infty|\sigma)$ to $P(r_n|\sigma)$, we exploit that the effective rate constants are independent of the choice of the dividing interface, $k_{\rm on}(\sigma) = k_{\rm on}(r_{\rm n})$, so that

$$k_{\rm D}(\sigma)(1 - P(\infty|\sigma)) = k_{\rm D}(r_{\rm n})(1 - P(\infty|r_{\rm n})), \tag{5.8}$$

where $1-P(\infty|\sigma) = k_{\rm a}(\sigma)/(k_{\rm a}(\sigma)+k_{\rm D}(\sigma))$ is the splitting probability for binding (c.f. Eq. 5.1). Factorizing $P(\infty|\sigma)$ as

$$P(\infty|\sigma) = P(\infty|r_{\rm n})P(r_{\rm n}|\sigma), \tag{5.9}$$

and simultaneously solving Eqs. 5.8 and 6.27 yields expressions for $P(\infty|r_{\rm n})$ and $P(\infty|\sigma)$. Substituting the expression of $P(\infty|\sigma)$ in Eq. 5.7 gives

$$k_{\rm a}(\sigma) = k_{\rm D}(\sigma)\frac{1 - P(r_{\rm n}|\sigma)}{P(r_{\rm n}|\sigma)}\left(1 - \frac{k_{\rm D}(\sigma)}{k_{\rm D}(r_{\rm n})}\right)^{-1} \tag{5.10}$$

Thus for a $r_{\rm n}$-interface chosen at a finite distance, we need to correct $k_{\rm a}$ by dividing through a factor $(1 - k_{\rm D}(\sigma)/k_{\rm D}(r_{\rm n}))$ The only unknown, $P(r_{\rm n}|\sigma)$, follows directly from the FFS/TIS simulation. This procedure can also be used to calculate the rate constants for particles interacting with anisotropic potentials, provided the position of the interface where these rates are measured is sufficiently far from contact, such that the orientational distributions on this interface are isotropic [1].

Now let us consider an interface, σ' between the σ-interface and $r_{\rm n}$-interface. The association rate $k_{\rm on}$ in terms of σ'- and $r_{\rm n}$-interfaces is given by [1]

$$k_{\rm on} = \frac{\left(1 - P(r_{\rm n}|\sigma')\right)k_{\rm D}(\sigma')}{1 - \Omega P(r_{\rm n}|\sigma')}, \tag{5.11}$$

where $\Omega = \sigma'/r_{\rm n}$. We showed in Ref. [1] that the value of $k_{\rm on}$ as a function of σ' becomes constant beyond $\sigma' > 3\sigma_{\rm an}$ for a non-specific interaction strength $\epsilon_{ns} = 10k_{\rm B}T$, and predicted that at this interface the orientational distributions of the particles become isotropic. We also in Ref. [1] showed that it is still possible to extract a meaningful value for the intrinsic rate constant at cross sections $\sigma < \sigma'$ by using the following expression:

$$k_a(\sigma) = \frac{(1 - P(r_n|\sigma'))k_D(\sigma')k_D(\sigma)}{P(r_n|\sigma')(k_D(\sigma') - k_D(\sigma)\Omega) + k_D(\sigma) - k_D(\sigma')}. \tag{5.12}$$

In this paper, we study how the interface at which the distributions become isotropic changes as a function of the strength of the non-specific attraction.

5.2.2. PARTICLE MODELS AND INTERACTION POTENTIALS

We employ two models of particles: (*i*) spherical particles interacting via an isotropic interaction potential between their centres of mass; (*ii*) spherical particles with an isotropic

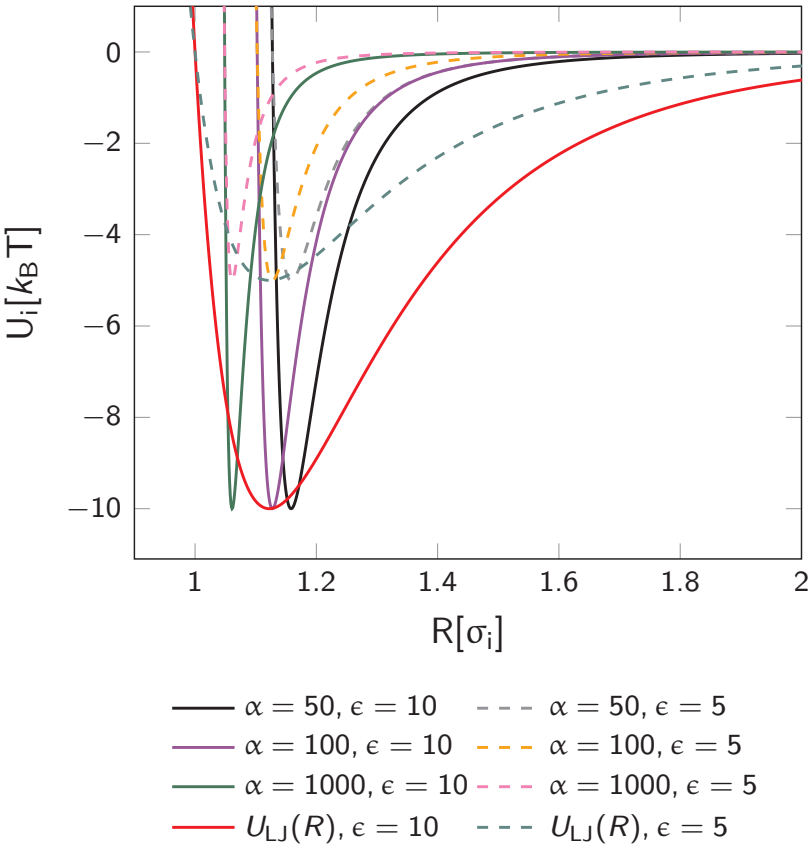
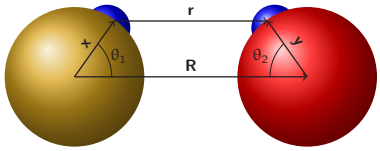
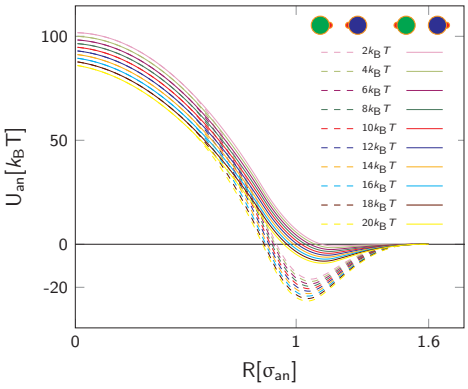


Figure 5.3: The isotropic potential, U_i (Eq. 5.13), as a function of the inter-particle distance R. The range and the interaction strength of the potential are set by α and ϵ respectively. A larger value of α results in a potential with a smaller range. The isotropic is potential plotted for three ranges ($\alpha = \{1000, 100, 50\}$) with each two different attraction strengths ($\epsilon = 5k_B T, 10k_B T$). The 12-6 Lennard-Jones potentials have also been plotted for comparison.The intrinsic and effective association and dissociation constants are evaluated for these six cases.

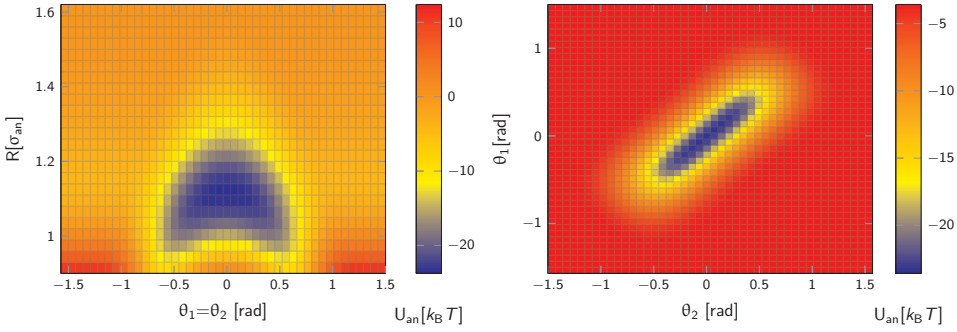
centre of mass interaction, dressed with one or more sticky spots on their surface called 'patches', which allow for highly directional, anisotropic interactions. In the following subsections we describe the isotropic and anisotropic potentials used to calculate the rate constants.



(a) Patchy particles



(b) Anisotropic potential



(c) Heat maps of the anisotropic potential

Figure 5.4: (a) Particles interacting via the anisotropic potential have sticky spots on the surface called 'patches'. These patches facilitate highly directional attractive interactions, which provides an effective model for proteins with binding sites. Particles interact via a combination of a repulsive force based on the centre-of-mass distance, R, between the particles, and a strong attractive force based on the distance between patches, r. In addition to this, particles experience a centre-of-mass distance based weak, non-specific, attractive force. This non-specific attraction facilitates a weakly bound state causing the particles not to immediately diffuse away once they unbind from the patches. This increases the probability of rebinding of the particles. (b) The anisotropic potential (Eq. 6.30) as a function of the inter-particle distance for aligned (dashed line) and misaligned (solid line) patches. The anisotropic potential is plotted for ten different non-specific attraction strengths ($\epsilon_{ns} = 2k_BT$, $4k_BT$, $6k_BT$, $8k_BT$, $10k_BT$, $12k_BT$, $14k_BT$, $16k_BT$, $18k_BT$, $20k_BT$). When the patches are aligned the distance between the patches is $r = R - \sigma_{an}$, and when misaligned $r = R + \sigma_{an}$, where σ_{an} is the length scale of the anisotropic potential and determines the particles' diameter. (c) Left: Heat map of the anisotropic potential as a function of distance R and the angle between the patch vector and the inter-particle vector, with $\theta_1 = \theta_2$ (angles as defined in (a)). Right: Heat map of the anisotropic potential as a function of θ_1 and θ_2, for a fixed $R = 1.1\sigma_{an}$. From these heat maps it follows that the specific attraction is strong only for a narrow range of orientations.

ISOTROPIC POTENTIAL

For the isotropic interaction we use the Lennard-Jones-inspired potential from Ref. [25], which allows easy control of the interaction range.

$$U_{\mathrm{i}}(R) = \frac{4\epsilon}{\alpha^2}\left(\left[\left(\frac{R}{\sigma_{\mathrm{i}}}\right)^2 - 1\right]^{-6} - \alpha\left[\left(\frac{R}{\sigma_{\mathrm{i}}}\right)^2 - 1\right]^{-3}\right), \tag{5.13}$$

where R is the centre-of-mass distance between the particles. This potential diverges asymptotically at $R = \sigma_i$, as opposed to the 12-6 LJ potential, which asymptotically diverges at $R = 0$. The strength of interaction is set by ϵ and the interaction range by α. Larger value of α means smaller interaction range. 5.3 shows the potential for three values of $\alpha = \{50, 100, 1000\}$ as a function of R, each plotted for two different interaction ranges ($\epsilon = 5k_{\mathrm{B}}T$ and $10k_{\mathrm{B}}T$). In our simulations, $\sigma_{\mathrm{i}} = 5$nm, roughly corresponding to a protein's diameter. For the sake of comparison, we also plot the standard 12-6 Lennard-Jones interaction, given by

$$U_{\mathrm{LJ}}(R) = 4\epsilon\left[\left(\frac{\sigma_{\mathrm{LJ}}}{R}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{R}\right)^{6}\right]. \tag{5.14}$$

Note that this standard LJ potential is indeed much longer ranged.

ANISOTROPIC POTENTIAL

Patchy particles (illustrated in Fig. 5.4a) model can proteins in an (idealized) coarse-grained way, where the patches represent the binding sites on the protein. Pairs of patchy particles, in our model, experience a strong attractive potential, $U_{\mathrm{s}}(r)$, over a narrow band of orientations (see Fig. 5.4c). 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 a repulsive potential, $U_{\mathrm{rep}}(R)$, which is a function of the center-of-mass distance, R. In addition, particles experience a weak, isotropic, non-specific attraction, $U_{\mathrm{ns}}(R)$. The total patch potential reads:

$$U_{\mathrm{an}}(R, r) = U_{\mathrm{s}}(r) + U_{\mathrm{rep}}(R) + U_{\mathrm{ns}}(R), \tag{5.15}$$

where $U_{\mathrm{s}}(r)$, $U_{\mathrm{rep}}(R)$ and $U_{\mathrm{ns}}(R)$ have the form

$$U_i(x) = \begin{cases} \epsilon_i\left(1 - a_i\left(\frac{x}{\sigma_{\mathrm{an}}}\right)^2\right) & \text{if } x < x_i^\star, \\ \epsilon_i b_i\left(\frac{x_i^c}{\sigma_{\mathrm{an}}} - \frac{x}{\sigma_{\mathrm{an}}}\right)^2 & \text{if } x_i^\star < x < x_i^c, \\ 0 & \text{otherwise,} \end{cases} \tag{5.16}$$

with $i = \{\mathrm{s}, \mathrm{rep}, \mathrm{ns}\}$ respectively. The overall strength ϵ_i, the length scale σ_{an}=5nm, 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 set the following parameters: $\epsilon_{\mathrm{s}} = 20k_{\mathrm{B}}T$, $a_{\mathrm{s}} = 20$ and $r^*_{\mathrm{att}} = 0.1\sigma_{\mathrm{an}}$, implying $b_{\mathrm{s}} = 5$ and $r^c_{\mathrm{s}} = 0.5\sigma_{\mathrm{an}}$; $\epsilon_{\mathrm{rep}} = 100k_{\mathrm{BT}}$, $a_{\mathrm{rep}} = 1$ and $R^*_{\mathrm{rep}} = 0.85\sigma_{\mathrm{an}}$, implying $b_{\mathrm{rep}} = 2.6036$ and $R^c_{\mathrm{rep}} = 1.1764\sigma_{\mathrm{an}}$; and $a_{\mathrm{ns}} = 1$

and $R^*_{ns} = 0.85\sigma_{an}$, implying $b_{ns} = 2.6036$ and $R^c_{ns} = 1.1764\sigma_{an}$. ϵ_{ns} is varied from $2k_BT$ to $20k_BT$ with steps of $2k_BT$. Fig. 5.4b shows the total potential as a function of R, when the patches are aligned ($r = R - \sigma_{an}$) and misaligned ($r = R + \sigma_{an}$). 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, $U_s = 0$ and the particles only experience the weak U_{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.

5.2.3. BROWNIAN DYNAMICS OF PATCHY PARTICLES

Propagation of the particle dynamics is done with Brownian dynamics, where each time step δt the particles position and orientation are updated based on the the instantaneous total force and torque acting on the particle. This total force/torque on the particle can be divided into three parts: $F_{tot} = F_v + F_p + F_r$. F_v is the viscous drag force, arising from the motion of the particle in a viscous solvent. F_p is the potential force due to interactions between the solute particles, as specified above. F_r is the random force, which models the interaction of the larger solute particles with smaller solvent particles.

In case of anisotropic interactions the total force and the torque is calculated, where the orientation of the particles are represented using quaternions. We use the Brownian dynamics integrator derived in Ref. [30] to simulate the rotational particle dynamics. Parameters that important in this integrator are the temperature T, the time step δt, and the translational and rotational friction coefficients $\gamma = \frac{k_BT}{D_t m}$ and $\Gamma = \frac{k_BT}{D_r M}$, respectively, where D_r and D_t are the translational and rotational diffusion coefficients, respectively, with k_B Boltzmann's constant. The mass of the particle m and the mass moment of inertia M are needed due to the formalism of the algorithm, but drop out during the integration, so that the dynamics will be not dependent on inertia. For more details see Ref. [30] and Refs. [23, 1, 24].

5.2.4. DISSOCIATION RATES BY FORWARD FLUX SAMPLING

As explained in Sec. 5.2.1, evaluation of the rate constants requires simulation of the dissociation reaction. Since dissociation is a rare event, brute force BD is very inefficient, and we have to use a rare event sampling technique such as TIS[27] or FFS. Here we use the 'direct-FFS' variant [29] to compute the rate constants. FFS is a simple, computationally efficient and inherently parallel algorithm to obtain good statistics of rare-event kinetics. At heart, FFS uses a series of interfaces between the bound and unbound state to calculate the transition path ensemble and calculate the corresponding transition rate. Trajectories starting in the bound state and reaching the unbound state are rare, but those starting at an interface and reaching the next interface is more common, if the interfaces are placed sufficiently close to each other. The interfaces are defined by a suitable order parameter λ: $(\lambda_0 ... \lambda_{n-1})$. FFS assumes that all trajectories from the bound to the unbound state should pass through all the interfaces in succession and that $\lambda_{i+1} > \lambda_i$ for all i. The order parameters used to define the interfaces in FFS are given below.

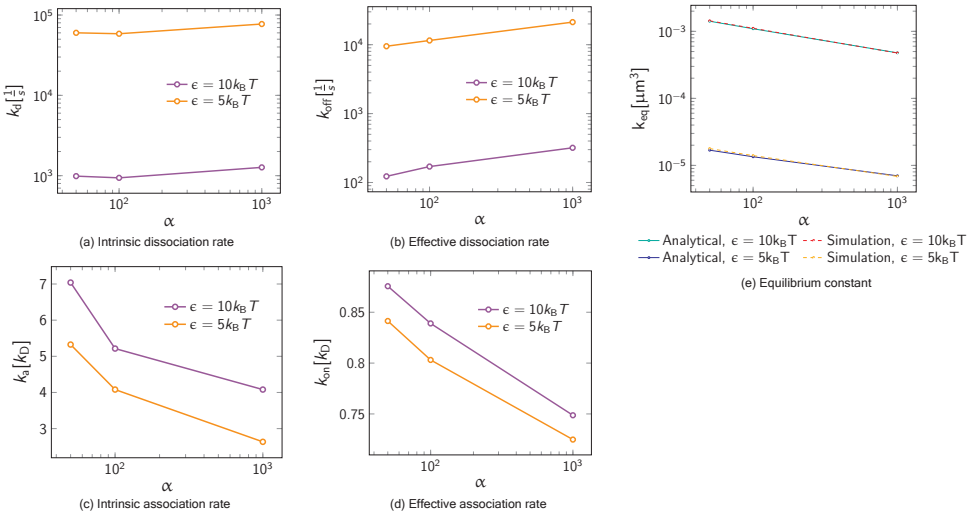


Figure 5.5: The intrinsic (a,c) and effective (b,d) rate constants and the equilibrium constants (e) as a function of the range of the isotropic potential (see Eq. 5.13), plotted for two different values the strength of the potential, $\epsilon = 5k_BT, 10k_BT$. The range of the potential is set by the parameter α, where a larger value of α results in a potential having a smaller range. The intrinsic rates are calculated at the σ-interface ($R = 1.6\sigma_i$), beyond which the value of the potential is truncated to zero. The values of k_d and k_{off} increase with decreasing range and increasing strength of the potential. The values of k_a and k_{on} decrease with decreasing range and increasing strength of the potential. In panel (e), the equilibrium constant is either calculated via $K_{eq} = k_a/k_d = k_{on}/k_{off}$, where $k_a, k_d, k_{on}, k_{off}$ are computed via the simulation technique presented in this paper, or analytically, by integrating the interaction potential, see Ref. [23].

5.2.5. SIMULATION DETAILS

The system specific parameters of the simulation are as follows: The time step $\delta t = 0.1$ns for the anisotropic potential and $\delta t = 10$ns for the isotropic potential, the mass of the particle is $m = 50$kDa, the mass moment of inertia $M = \frac{8}{15} m\sigma_{an}^2$ the translational and rotational diffusion constants, are $D_t = 1\mu m^2/s$ and $D_r = 1.6 \times 10^7 rad^2/s$ for all particles, the translational and rotational friction coefficients are $\gamma = \frac{k_B T}{D_t m}$ and $\Gamma = \frac{k_B T}{D_r M}$ respectively, where $k_B = 1.38 \times 10^{-23} JK^{-1}$ is the Boltzmann constant and $T = 300$K is the temperature of the system.

For the isotropic potential, the order parameters that define interfaces in the FFS simulation, is based on the distance R between the centres-of-mass of the particles. The first interface (r_0-interface), which defines the bound state is at $R = 1.2\sigma_i$. Successive interfaces are placed at $R = 1.3, 1.4, 1.5$ and $1.6\sigma_i$. At $1.6\sigma_i$, the isotropic potential is truncated to zero, and this interface is the σ-interface. Beyond the cutoff there is only one interface at $R = 2.0\sigma_i$: this is the r_n-interface, which is used to compute the escape probability $P(\infty|\sigma)$.

In case of the anisotropic potential, the interfaces are defined in terms of energy,

$U_{\mathrm{an}}(R,r)$ until the cut-off of the potential, beyond which they are based on the inter-particle distance, R. The first interface (r_0-interface), defining the bound state is placed at $18k_{\mathrm{B}}T$. The successive interfaces are located at 15, 10, $5k_{\mathrm{B}}T$. The σ-interface at the cut-off of the potential is defined by zero energy and $R = 1.6\sigma_{\mathrm{an}}$. Beyond the cut-off of the potential, interfaces are placed at R=1.7, 1.9, 2.1, 2.3, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5 and finally the r_{n}-interface is placed at $R = 7\sigma_{\mathrm{an}}$.

5.3. RESULTS

We calculated the rate constants by systematically varying the strength and the range of interaction for the isotropic potential and the non-specific attraction strength for the anisotropic potential. In the latter case of the anisotropic potential we also determine, as a function of the non-specific attraction strength, the interface beyond which the expressions of these rate constants are valid.

5.3.1. RATE CONSTANTS FOR THE ISOTROPIC POTENTIAL

We calculate the intrinsic and effective, association and dissociation rates as detailed in 5.2.1 for different values of the range (α) and the strength (ϵ) of the isotropic potential. The r_0, σ and r_{n} interfaces are at $R = 1.2\sigma_{\mathrm{i}}$, $R = 1.6\sigma_{\mathrm{i}}$ and $R = 2.0\sigma_{\mathrm{i}}$ respectively.

Figs. 5.5a and 5.5b plot the dissociation rates k_{d} and k_{off}, respectively, for two different values of interaction strength, ϵ, as a function of α, which sets the range of the potential (larger α results in a potential with a smaller range). Both k_{d} and k_{off} increase marginally with decreasing range, but increase significantly when the interaction strength is halved. Decreasing range or the strength facilitates easier unbinding of the particles and hence increases the dissociation rate constants. Hence, the strength of the potential has a large influence on the dissociation rates.

Figs. 5.5c and 5.5d show k_{a} and k_{on}, respectively, as a function of α, for two different values of interaction strength, ϵ. Here, k_{a} decreases by roughly a factor of two with decreasing range and only marginally when the interaction strength is halved. The effective rate constant k_{on} decreases only marginally with decreasing range and decrease even less when the interaction strength is halved. The range and the strength of the potential, have thus a very small influence on the association rates.

Figs. 5.5e shows the equilibrium constant computed from $K_{\mathrm{eq}} = k_a/k_d = k_{\mathrm{on}}/k_{\mathrm{off}}$ as a function of α, for two different values of interaction strength, ϵ. For comparison we included the values obtained from the analytical expression $K_{\mathrm{eq}} = K_D^{-1} = 4\pi\int_0^{r_p} r^2 e^{-\beta V(r)}\,dr$. The simulations agree extremely well with the analytical expression.

5.3.2. ORIENTATIONAL DISTRIBUTION FOR THE ANISOTROPIC POTENTIAL

First, we determine how far from the σ-interface the orientational distributions of the particles become isotropic, as a function of the non-specific attraction strength, ϵ_{ns}. We determine this interface using two approaches: (i) Constructing the orientational distributions of the particles and checking at which interface these distributions become isotropic.(ii) Identifying the interface at which the value of k_{on}, calculated from Eq. 5.11, converges to a constant value [1].

The orientational distributions are constructed from an extensive brute force BD

simulation of two particles initially in the bound state. As the simulation progresses, the particles move away from each other. The position and orientation of the particles are recorded each time they cross one of the FFS interfaces at position r beyond the potential cut-off distance σ. The orientation is monitored via two parameters: 1) the angle θ between the inter-particle vectors at the initial bound state and the particles' current position, 2) the angle α between the patch vectors of the two particles. These two angles are illustrated at the top of 5.6. The remainder of Fig. 6 plots the probability distributions for the angle θ (left column) and of the angle α (right column) for five interfaces $R = 1.7, 1.9, 2.1, 2.3\sigma_{\text{an}}$. The interface where the angular distributions becomes isotropic is indicated in grey; a distribution is considered to be isotropic when. The isotropic distribution is plotted in orange for the sake of comparison. The peak in the θ distributions is shifted to lower angles for relatively small values of non-specific attraction strength ϵ_{ns} at interfaces near the bound state, due to a strong correlation with the initial orientation. At interfaces further away, due to diffusion and the truncation of the potential, particles are less correlated to the initial orientation and position, and are eventually isotropically distributed. However, for larger values of the non-specific attraction strength, ϵ_{ns}, the distribution of θ becomes isotropic already at interfaces close to the bound state. This loss of correlation is caused by particles lingering longer in the non-specifically bound state and not immediately diffusing away. The probability distribution of α in the right column of 5.6 becomes isotropic at interfaces closer to the bound state, when compared to the distribution of θ for the same non-specific attraction strength, ϵ_{ns}. We also observe that with increasing ϵ_{ns} the shift in the interface where the orientational distributions α become isotropic does not shift as drastically as for the distributions of θ.

5.3.3. RATE CONSTANTS FOR THE ANISOTROPIC POTENTIAL

Next, we study the behaviour of the rate constants by increasing the non-specific attraction strength ϵ_{ns}. We calculate the rate constants as explained in Sec. 5.2.1. As discussed above, it is important to determine the interface at which the orientational distributions become isotropic, so that the expressions of the association rate constants are valid for anisotropic potentials. This can either be done by plotting the orientational distributions (see previous section and 5.6), or by calculating $k_{\text{on}}(\sigma')$ and determine the value of σ' for which the effective association rate becomes independent of σ', and reaches a constant. To this end, we computed $k_{\text{on}}(\sigma')$ as a function of σ' using Eq. 5.11, for ten different values of the non-specific interaction strength ϵ_{ns}, see Fig. 5.7a. We observe that as ϵ_{ns} increases, the value of k_{on} increases. We also notice that for interfaces close to the bound state, Eq. 5.11 predicts incorrect values for k_{on}. However, as we move further away the orientational distributions become isotropic and k_{on} converges to the correct value. To quantitatively determine the interface where the value of k_{on} converges to the correct value, we calculate the relative error in k_{on}. The relative error in k_{on} is given by $\frac{k_{\text{on}}(\sigma')}{k_{\text{on}}(5.5\sigma_{\text{an}})} - 1$, where $k_{\text{on}}(\sigma')$ is the value of at a given interface σ' and $k_{\text{on}}(5.5\sigma_{\text{an}})$ is the value at the interface at $R = 5.5\sigma_{\text{an}}$. Fig. 5.7b shows this relative error plotted as a function of the σ'-interface for ten different values of ϵ_{ns}. We assume that if the relative error is smaller than 0.1%, the value of k_{on} has converged to the correct value and at this point we predict that the orientational distributions become isotropic. The interface value where $k_{\text{on}}(\sigma')$ converges is plotted as a function of ϵ_{ns} in Fig. 5.7c. For a smaller value

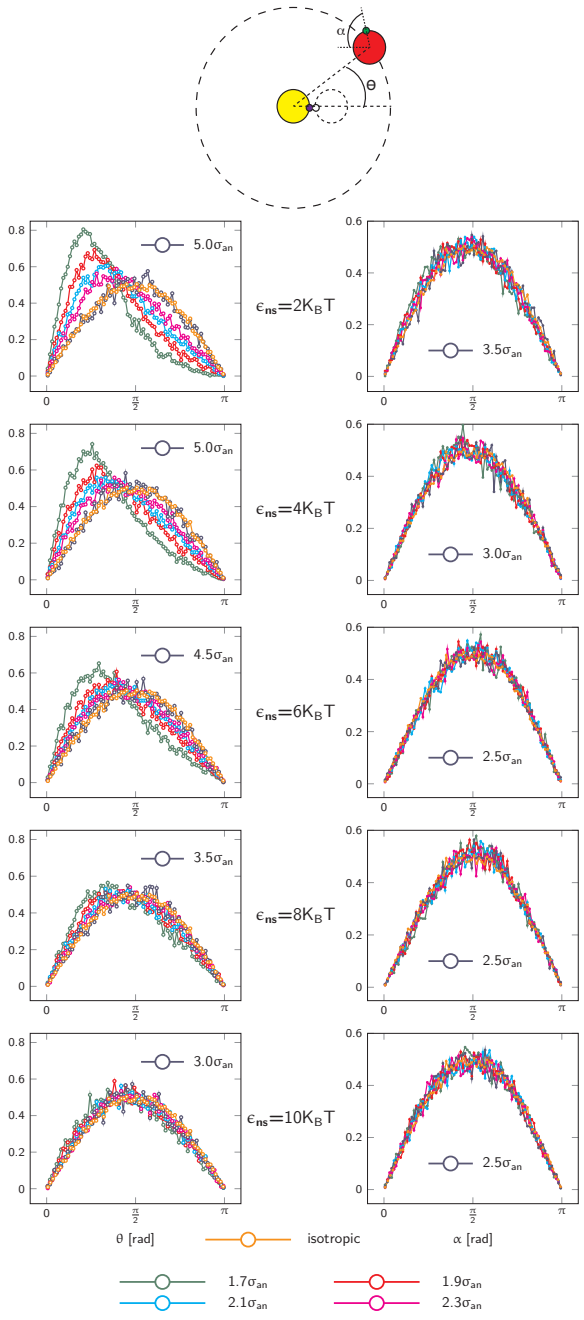
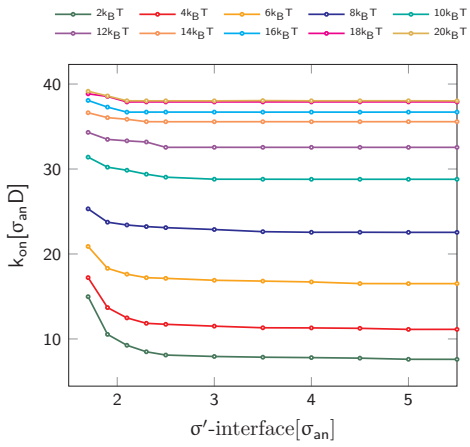
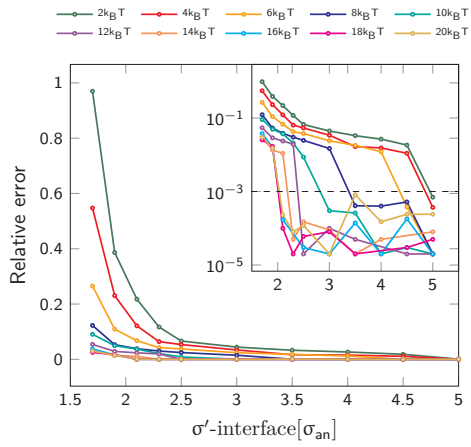


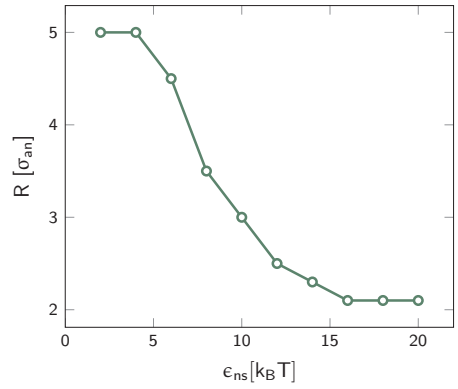
Figure 5.6: The distributions of angles θ (left) and α (right) at different interfaces. θ is the angle between the initial (bound state) centre-of-mass vector and the centre-of-mass vector when the particle reaches the interface. α is the angle between the patch vector of the particle at the initial bound state and the patch vector when the particle reaches the interface. These angles are illustrated in the cartoon at the top of the figure. We see that the distribution of α becomes isotropic faster than the distribution of θ for any given ϵ_{ns}.



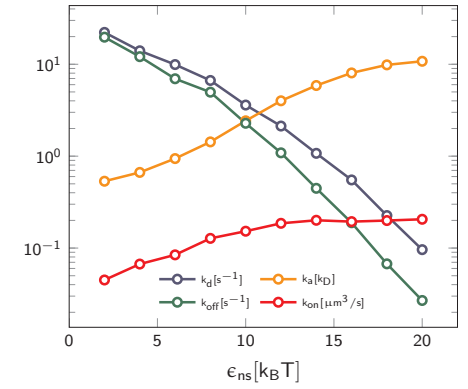
(a) k_{on}



(b) Relative error in k_{on}



(c) Interface where k_{on} becomes constant



(d) k_a, k_d, k_{on} and k_{off}

Figure 5.7: The rates of association and dissociation for the anisotropic interaction potential.(a) k_{on} calculated from Eq. 5.11 as a function of the position of the σ'-interface, plotted for ten different values of the non-specific attraction strength, ϵ_{ns}. The value of k_{on} increases with increasing ϵ_{ns}. The plot shows k_{on} converging to the correct value as we move σ' further from the bound state. This is because Eq. 5.11 is valid only if the orientational distributions are isotropic in the interface where we measure k_{on} (b) The relative error in evaluating k_{on} as function of σ'. At the interface where the value of k_{on} becomes a constant, the orientational distributions become isotropic. To determine this interface, we calculated the relative error $\frac{k_{on}(\sigma')}{k_{on}(5.5\sigma_{an})} - 1$. If this relative error is below 0.01%, we conclude k_{on} has converged. The inset shows the same plot in a semi-log scale. The dashed line denotes 0.1% relative error. (c) The interface at which the value of k_{on} becomes a constant as a function of the non-specific attraction strength ϵ_{ns}. The orientational distributions become isotropic at smaller distances for higher ϵ_{ns}, because the particles spend more time near each other and have more time to decorrelate from the bound state configuration. (d) k_a, k_d, k_{on} and k_{off} as a function of the non-specific attraction strength. k_{off} and k_d decrease significantly with increasing ϵ_{ns}. The k_{on} and k_a on the other hand, increase initially with increasing non-specific attraction strength, but later levels off. The increase in the association rates for a potential with ten times stronger ϵ_{ns} is one order of magnitude.

of ϵ_{ns} particles need to move further away from the bound state for the value of k_{on} to converge. For a larger value of ϵ_{ns}, k_{on} converges at smaller distances since the particles stay longer in the nonspecific state, before diffusing away. We also predict that when k_{on} converges the orientational distributions become isotropic. Indeed, the interface where the distribution of θ becomes isotropic as shown in 5.6, matches with the the interface where the value of k_{on} converges as shown in Fig. 5.7c.

Finally, Fig. 5.7d shows the behaviour of all the four converged rate constants as a function of ϵ_{ns}, evaluated at a cross section $\sigma = 1.6\sigma_{an}$. The intrinsic rates k_a were computed using Eq. 5.12. The dissociation rate constants k_d and k_{off} decrease with increasing ϵ_{ns}. In contrast, k_a and k_{on} increase initially with increasing ϵ_{ns} because the non-specific attraction promotes specific binding, as particles stay in each others vicinity. The increase in the association rates is not dramatic with increasing non-specific attraction strength: about an order of magnitude for a ten times increase in non-specific attraction strength. This behavior was also found in Ref. [31]. Moreover, beyond a limiting value of ϵ_{ns}, the association rates reach a plateau. This is clearly caused by a flattening of the intrinsic association rate constant. This raises the question, whether the intrinsic rate constant can ever become very high, as is assumed in many diffusion influenced rate theories.

5.4. DISCUSSION AND CONCLUSIONS

In this work we have evaluated intrinsic and effective rate constants using the explicit microscopic expressions that we presented in Ref. [1]. Knowledge of these rates is important to study the microscopic dynamics of reaction-diffusion systems. Furthermore, to construct Markov State Models of reaction-diffusion systems, these rates are essential input parameters. Here we studied these rates in the context of an isotropic and an anisotropic potential. In the case of the isotropic potential we evaluated these rates as a function of the range and the strength of the potential. We observed that the dissociation rates increase slightly as the interaction range is decreased, yet increases drastically when the strength of the potential is halved. On the other hand, the association rates decrease with decreasing range. The change in the effective association rate constants is not significant.

In case of the anisotropic potential, we studied the behaviour of these rates as a function of the non-specific isotropic attraction. In addition to the bound and unbound states, this non-specific attraction facilitates a third non-specifically bound state in which the particles are attracted to each other, but not strongly bound to a particular patch. This increases the possibility of the particles to realign and bind at a specific patch, before diffusing away. The expression for the rates that we derived holds only if the orientational distributions of the particles are isotropic at the cross-section where the rates are measured. We determined this interface directly from two angular distributions of the particles, and by calculating the effective association rate as a function of the interface and identifying where this rate constant is converged. The interfaces obtained from both methods agree with each other. Knowledge of the location of this interface becomes important when we use the computed intrinsic rates to construct a Markov State Model, and combine them with the mesoscopic Greens Function Reaction Dynamics method

within a multi-scale scheme [23, 24].

The simulations also reveal that the dissociation rate decreases drastically when the non-specific isotropic attraction strength is increased (see Fig. 5.7). This is because non-specific binding increases the likelihood that particles that have just dissociated rebind instead of diffusing away. The association rate increases with the non-specific isotropic attraction strength. Non-specific binding keeps the particles that have diffused toward one another in close proximity, given them time to align their patches and bind specifically. This effect is akin to the antenna effect in the binding of transcription factors to their specific sites on the DNA [32]: non-specific binding increases the effective cross-section for the binding of proteins to their specific site. Our simulations also show that the effect of non-specific binding on the association rate is much weaker than that on the dissociation rate: while the effective and intrinsic dissociation rates decrease by more than two orders of magnitude when the non-specific interaction strength ϵ_{ns} is increased from $2k_BT$ to $10k_BT$, the effective association rate increases by less than an order of magnitude [31]. Moreover, while the effective dissociation rate continues to decrease as ϵ_{ns} is increased, the effective association rate saturates. This is because the overall association rate becomes increasingly limited by diffusion. Indeed, the intrinsic association rate continues to increase with ϵ_{ns}, although a close inspection of Fig. 5.7 shows that this rise levels off too—also the intrinsic association ultimately becomes limited by diffusion.

Many reactions are believed to be diffusion limited. This means that the intrinsic rate k_{a} is much higher than the diffusion-limited arrival rate k_{D}. Yet, how much larger the intrinsic rate can be, has, to our knowledge, not been systematically addressed before. Addressing this question, it should first be realized that the values of both k_{a} and k_{D} depend on the choice of the cross section σ: the larger σ, the larger the diffusion-limited arrival rate k_{D} and the lower the intrinsic rate k_{a}. However, a natural choice for σ is the distance as given by the effective physical size of the particles. After all, the intrinsic rate is typically interpreted as the rate at which the particles react given that they are in contact. Moreover, a small cross section also facilitates the modelling of many-body reaction-diffusion systems—the larger the cross section, the more often three-body (and higher) interactions have to be taken into account. Yet, at the microscopic scale there is no unique definition for what the effective physical size is: one choice is the distance where the particles start to repel each other; another is the range of the interaction potential. However, choosing the first option would violate the basic assumption made in Eq. 5.1, namely that the cross section has to be chosen beyond the potential cutoff. Fortunately, for proteins interacting via isotropic potentials, the range of the potential is typically short, and hence the ambiguity in the definition is not critical. The natural choice for σ is therefore the range of the interaction potential, or more specifically its cutoff. This also means the intrinsic association rate k_{a} cannot be made arbitrarily large, but has an upper limit. This maximally achievable value of k_a will also depend on the shape of the isotropic potential. For instance, replacing the potential with a square well potential, of the same depth, while keeping the equilibrium constant K_{eq} fixed, moves the potential cutoff to lower values, with a corresponding higher k_{a}. For the isotropic Lennard Jones-based systems considered in this work, the intrinsic rate k_{a}, computed for a cross section σ given by the cutoff of the potential, is about a factor 1-10 higher than the diffusion-limited rate k_{D}. Thus, while these systems are in the diffusion-limited

regime, they are still influenced by the intrinsic association rate k_a.

For particles interacting via anisotropic potentials, the interaction range is short for a given orientation of the particles. However, the distance beyond which the orientational distribution of the dissociating particles has become isotropic, is much longer (see 5.6). It then follows from detailed balance that also the distribution of associating particles (more specifically, the distribution of trajectories that start in the unbound state and end in the bound state) becomes isotropic only beyond this distance. In this regime, one can truly speak of one well-defined intrinsic rate $k_a(\sigma)$, one diffusion-limited arrival rate $k_D(\sigma)$, and one cross-section σ, independent of the orientation of the particles. However, as our earlier work [1] and that of Northrup et al. [33, 3] show, it is still possible to talk about intrinsic association rate constants for cross section values that are smaller than the distance where the orientational distribution of association and dissociation trajectories becomes isotropic. These intrinsic rate constants should then be viewed as an average over all orientational dependent intrinsic rates. Indeed, as long as σ is chosen beyond the cut-off of the potential, our approach, via Eq. 5.12, does also make it possible to obtain k_a and k_D in this regime where the distribution of association/dissociation trajectories has not become isotropic yet.

Since the ratio k_a/k_D increases with decreasing σ, to address the maximal value of k_a/k_D, we thus compute k_a and k_D for the smallest possible choice of σ, which is again the cut-off of the potential. This ratio k_a/k_D depends on the strength of the non-specific attraction, see Fig. 5.7. When the non-specific attraction is weak, $k_a \approx 0.7k_D$, which means that association is reaction limited. This is because of the strong anisotropy of the interaction potential: the particles can only bind when their patches are properly aligned with each other, limiting the binding probability (of course, this reaction itself is also a diffusion process). Yet, the figure also shows that when ϵ_{ns} is increased, the intrinsic association rate increases. But it does not increase indefinitely. Instead k_a levels off, at a value that is on the order of 10 k_D (see Fig. 5.7d). With $k_a \approx 10k_D$, the effective association rate is dominated by the diffusion-limited arrival rate, yet still influenced by the intrinsic association rate.

Finally, it would also be of interest to study the effect of the shape of the potential on the values of the rates constant. For example, it would be of interest to compare the rates computed here for the modified LJ potential to those of a square-well potential. This comparison would then have to be performed on the footing of equal well-depth and equal range of the potential, since this guarantees that the phase behaviour and the equilibrium constants are very similar. We leave this for future work.

5.5. ACKNOWLEDGEMENTS

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.

BIBLIOGRAPHY

[1] Vijaykumar A, Bolhuis PG, ten Wolde PR (2016) The intrinsic rate constants in diffusion-influenced reactions. *Faraday Discussions* 195:421–441.

[2] Agmon N, Szabo A (1990) Theory of reversible diffusion-influenced reactions. *The Journal of Chemical Physics* 92:5270.

[3] Northrup SH, Erickson HP (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proceedings of the National Academy of Sciences* 89:3338–3342.

[4] Zhou HX (1998) Comparison of three Brownian-dynamics algorithms for calculating rate constants of diffusion-influenced reactions. *The Journal of Chemical Physics* 108:8139.

[5] van Zon JS, Morelli MJ, Tanase-Nicola S, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expresssion. *Biophysical Journal* 91:4350.

[6] Kaizu K, et al. (2014) The berg-purcell limit revisited. *Biophysical journal* 106:976–985.

[7] Takahashi K, Tănase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America* 107:2473–8.

[8] Gopich IV, Szabo A (2013) Diffusion modifies the connectivity of kinetic schemes for multisite binding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 110:19784–19789.

[9] Aokia K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:12675–12680.

[10] Berg HC, Purcell EM (1977) Physics of chemoreception. *Biophysical Journal* 20:193.

[11] Bialek W, Setayeshgar S (2005) Physical limits to biochemical signaling. *Proceedings of the National Academy of Sciences USA* 102:10040.

[12] Wang K, Rappel WJ, Kerr R, Levine H (2007) Quantifying noise levels in intercellular signals. *Physical Review E* 75:061905.

[13] Berezhkovskii AM, Szabo A (2013) Effect of ligand diffusion on occupancy fluctuations of cell-surface receptors. *The Journal of Chemical Physics* 139:121910.

[14] ten Wolde PR, Becker NB, Ouldridge TE, Mugler A (2016) Fundamental Limits to Cellular SensingPieter Rein ten Wolde, Nils B. Becker, Thomas E. Ouldridge & Andrew Mugler. *Journal of Statistical Physics* 162:1395–1424.

[15] Tkačik G, Bialek W (2009) Diffusion, dimensionality, and noise in transcriptional regulation. *Physical Review E* 79:051901.

[16] Paijmans J, ten Wolde PR (2014) Lower bound on the precision of transcriptional regulation and why facilitated diffusion can reduce noise in gene expression. *Physical Review E* 90:032708.

[17] Bicknell BA, Dayan P, Goodhill GJ (1) The limits of chemosensation vary across dimensions. *Nature Communications* 6:1–8.

[18] van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *The Journal of chemical physics* 123:234910.

[19] Morelli MJ, ten Wolde PR (2008) Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network. *The Journal of Chemical Physics* 129:054112.

[20] Lipková J, Zygalakis KC, Chapman SJ, Erban R (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. *J. Appl. Math.* 71:714–730.

[21] Johnson ME, Hummer G (2014) Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. *Phys. Rev. X* 4:031037.

[22] Klein HCR, Schwarz US (2014) Studying protein assembly with reversible Brownian dynamics of patchy particles. *The Journal of Chemical Physics* 140:184112.

[23] Vijaykumar A, Bolhuis PG, ten Wolde PR (2015) Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations. *The Journal of Chemical Physics* 143:214102.

[24] Vijaykumar A, Ouldridge TE, ten Wolde PR, Bolhuis PG (2017) Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics. *The Journal of Chemical Physics* 146:114106.

[25] ten Wolde PR, Frenkel D (1997) Enhancement of protein crystal nucleation by critical density fluctuations. *Science* 277:1975–1978.

[26] van Zon JS, ten Wolde PR (2005) Simulating Biochemical Networks at the Particle Level and in Time and Space: Green's Function Reaction Dynamics. *Physical Review Letters* 94:1–4.

[27] van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. *The Journal of Chemical Physics* 118:7762.

[28] Allen RJ, Warren PB, ten Wolde PR (2005) Sampling Rare Switching Events in Biochemical Networks. *Physical Review Letters* 94:018104.

[29] Allen RJ, Valeriani C, Rein ten Wolde P (2009) Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter* 21:463102.

[30] Davidchack R, Ouldridge T, Tretyakov M (2015) New langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* 142.

[31] Newton AC, Groenewold J, Kegel WK, Bolhuis PG (2017) *submitted*.

[32] Hu T, Grosberg AY, Shklovskii BI (2006) How Proteins Search for Their Specific Sites on DNA: The Role of DNA Conformation. *Biophysical Journal* 90:2731–2744.

[33] Northrup SH, Allison SA, McCammon JA (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80:1517.

6

TRANSITION RATES FOR PROTEINS ASSOCIATING TO SUBSTRATES WITH MULTIPLE BINDING SITES USING A NOVEL GENERIC FORWARD FLUX SAMPLING EXPRESSION

Intrinsic and effective rates are key quantities to predict the response of a biochemical system. In our previous paper [1], we derived microscopic expressions to calculate the intrinsic and effective rate constants, using a single simulation of the dissociation reaction. However, these expressions are valid when the substrate has one binding patch. If the substrate has multiple binding sites, a bound enzyme can, besides dissociating into the bulk, also hop to another binding site. In this paper, we compute the association, dissociation, and hopping rates as a function of the distance between the binding sites and the rotational diffusion constant. We also determine the effect of blocking of one of these binding sites on the rates. We first derive a new generic expression to calculate transition rates using Forward Flux Sampling, when the states are not necessarily separated by all the interfaces. This is necessary to calculate transition rates between multiple states. Finally, we illustrate this approach by computing these rates for a system in which an enzyme binds to a substrate with two binding sites, and we show how the rates depend on the distance between the binding sites and the orientational diffusion constant of the particles.

6.1. INTRODUCTION

Unimolecular and bimolecular reactions are the building blocks of many complex processes in biology, chemistry, and soft condensed matter. When studying such molecular interactions, it is natural to use the equilibrium constants of the reactions to predict the direction of the reaction from the equilibrium concentrations of the reactants and products. The knowledge of equilibrium constants is imperative to evaluate the relevance of biological reactions like DNA-protein binding, receptor-ligand binding, enzyme-substrate binding. However, knowledge of the equilibrium constant alone is not sufficient to study the kinetics of these processes. The equilibrium constant is the ratio of the rate at which the reactants bind to the rate at which they unbind. For the same value of the equilibrium constant, the binding an unbinding rates can be very different. Binding and unbinding rates also determine the response time of living cells, which is vital for the fitness of the organism in fluctuating environments. Knowledge of the individual rates is also important in the field of drug development, for example, it is important to know how long the drug is bound to its target to set an optimal dosage and to improve drug efficacy [2]. Also discrimination by our immune system is believed to depend on the dissociation rate, rather than the equilibrium constant.

Many substrates (proteins, DNA) have multiple binding sites, to which enzymes can bind. When bound to a binding site of the substrate, the enzyme can, for instance, phosphorylate the substrate. Subsequently, the enzyme can dissociate from this site, and diffuse into the bulk, and later bind to the other (or the same) site. This is called a distributive mechanism. Alternatively, the enzyme can hop between the two binding sites, without dissociating and moving to the bulk, in a processive mechanism. The response of such a network can change drastically based on the mechanism of the rebinding to the other site [3]. To predict this behavior requires computation of the microscopic rates of hopping, association and dissociation. These rates are also dependent on environmental conditions. For instance, In a crowded environment, the rotational diffusion constant of the proteins can be varied independent of the translational diffusion constant by varying the nature and the concentration of the crowders [4].

In a previous paper [1], we derived explicit microscopic expressions for the intrinsic and effective rate constants and describe a technique to compute these rates using rare-event simulation techniques. Using these expressions, we can not only calculate the binding and unbinding rates of isotropic particles, but also those of anisotropic particles, with orientational dynamics. The intrinsic rate constants are important to model biological reactions, and serve as input parameters for multi-scale techniques to simulate reaction-diffusion systems over multiple length and time scales[5, 6]. The effective rates are of interest to experimentalists since typically in experiments the effective rates are measured. The strength and novelty of our technique is to calculate all these rates with a single simulation of the dissociation reaction using a rare-event technique like forward flux sampling or transition interface sampling [4, 7, 8].

In this paper, we first derive the expression to calculate the transition rate using forward flux sampling, in the case where the initial and final states are not separated by all the intermediate interfaces. This is necessary to calculate the hopping rates between the multiple binding sites. Next we derive expressions for the binding and unbinding rates in the case that a substrate has two binding sites. Finally we compute and compare

the binding and unbinding rates in five cases: (1) the substrate has one binding site, (2) the substrate has two binding sites, but we do not distinguish between them, (3) the substrate has two binding sites, and we differentiate between them, (4) the substrate has two binding sites and one of them is blocked. In all the cases the enzyme has one binding site.

6.2. METHODS

6.2.1. RATE CONSTANTS FROM FORWARD FLUX SAMPLING WHEN TWO STATES ARE NOT SEPARATED BY ALL INTERFACES

Dissociation of anisotropic particles is often a rare event. In order to simulate such rare event kinetics with minimal computational effort, an efficient rare event simulation technique such as Forward Flux Sampling(FFS) [9, 10] or Transition Interface Sampling(TIS) [11] is necessary. Here, we use FFS to simulate the dissociation reaction.

The crux of FFS is to drive the system from one state to another in a ratchet-like manner by capitalizing on those fluctuations that happen to move the system in the right direction. To capitalize on these fluctuations, FFS uses a series of interfaces between the initial and final state. These interfaces make it possible to store configurations along trajectories that have progressed in the direction of the final state. While FFS is typically employed for computing transition rates between two states, here we present a new expression, which may prove useful for computing transition rates between multiple states.

Fig. 6.1 illustrates the scheme for a scenario of three metastable states: A, B, and U. These states are defined in terms of an order parameter λ. Here, we are interested in the transition rates k_{AB} and k_{AU}. While the expressions that we will derive below are generic, and can be generalized to any system consisting of multiple metastable states, it is illuminating to consider the concrete scenario in which state A corresponds to an enzyme molecule that is bound to patch A of a substrate molecule, state B as the enzyme molecule being bound to patch B of the substrate molecule, and state U as the state in which the enzyme and substrate molecules are unbound. The dissociation rate k_{AU} is then defined as the rate of dissociating from patch A into the unbound state U while not visiting state B; the ensemble of transition paths that corresponds to k_{AU} thus contains trajectories that start in A and end in U, yet do not visit B. In contrast, the hopping rate k_{AB} is defined as the rate at which the enzyme molecule dissociates from patch A and then diffuses to and rebinds to patch B. Importantly, the transition path ensemble that corresponds to this hopping rate contains not only trajectories that directly go from A to B, yet also contains trajectories that have significantly progressed in the direction of U before arriving in B. Here, we will derive the expressions that make it possible to compute the transition rates k_{AB} and k_{AU} in an FFS simulation.

As Fig. 6.1 illustrates, the interfaces $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}, \lambda_n$ are defined such that all trajectories of the transition path ensemble corresponding to k_{AU} necessarily cross all interfaces $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}, \lambda_n$. As a result, the expression for k_{AU} is based on the conven-

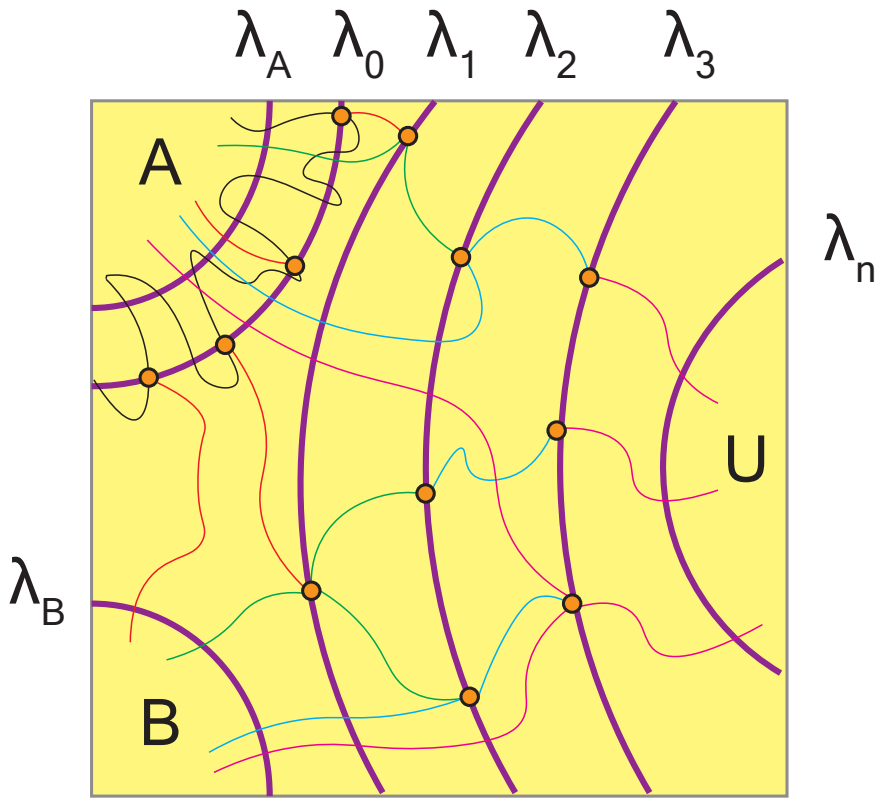


Figure 6.1: Illustration of the possible trajectories starting from a state A and ending either in a state B or a state U and the interfaces $\lambda_0 ... \lambda_{n-1}$ used in the FFS simulation. A trajectory starting at A and terminating at B, need not pass through all these interfaces, but a trajectory starting at A and terminating at U has to pass through all the interfaces.

tional TIS expression used also in FFS [11, 9]:

$$k_{AU} = \Phi_0 \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i). \tag{6.1}$$

Here, λ_i define the intermediate interfaces between state A and state U, as illustrated in Fig. 6.1. The quantity Φ_0 is the flux of trajectories that start in A and then cross interface λ_0, while $P(\lambda_{i+1}|\lambda_i)$ is the conditional probability that a trajectory that comes from A and crosses λ_i for the first time will subsequently reach λ_{i+1} instead of returning to A or progressing to B. In an FFS simulation, one thus first performs a brute-force simulation in state A; this makes it possible to not only compute the flux Φ_0 through the first interface λ_0, but also to generate an ensemble of points at λ_0. In the next step, one then randomly picks a configuration from this ensemble of points at λ_0, launches and propagates a trajectory from this configuration until it either arrives at λ_1, returns to A,

or arrives in B.; by iterating this a number of times, one obtains not only an ensemble of configurations at λ_1, but also $P(\lambda_1|\lambda_0)$ as the fraction of trajectories that reaches λ_1. This procedure is then repeated for all the subsequent interfaces, yielding $P(\lambda_{i+1}|\lambda_i)$ for all interfaces λ_i.

In contrast to the trajectories of the *AU* transition path ensemble, the trajectories of the *AB* path ensemble do not necessarily cross all interfaces $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}, \lambda_n$. Some paths directly go from A to B, while other trajectories cross λ_1 and perhaps even $\lambda_i > \lambda_1$, before proceeding to B. All these excursions must be accounted for to calculate the transition rate from A to B. This means that Eq. 6.1 cannot be used to compute k_{AB}. A general expression for the transition rate between two states, where the trajectories start in A and end in B yet do not necessarily cross all intermediate interfaces, is given by

$$\begin{aligned} k_{AB} = \Phi_0 \Big[& P(\lambda_\mathrm{B}|\lambda_0) + P(\lambda_1|\lambda_0)P(\lambda_\mathrm{B}|\lambda_1) + \\ & + P(\lambda_1|\lambda_0)P(\lambda_2|\lambda_1)P(\lambda_\mathrm{B}|\lambda_2) + \cdots + \\ & + P(\lambda_1|\lambda_0)\ldots P(\lambda_{\mathrm{n}-1}|\lambda_{\mathrm{n}-2})P(\lambda_\mathrm{n}|\lambda_{\mathrm{n}-1})P(\lambda_\mathrm{B}|\lambda_\mathrm{n}) \Big] \\ & = \Phi_0 \sum_{i=1}^{n} P(\lambda_\mathrm{B}|\lambda_i)P(\lambda_i|\lambda_0) \end{aligned} \tag{6.2}$$

Here, $P(\lambda_B|\lambda_i)$ is the probability that a trajectory that is launched at interface λ_i arrives in B before reaching either A or λ_{i+1}, while $P(\lambda_{i+1}|\lambda_i)$ is, as before, the probability that a trajectory that is fired at λ_i arrives at λ_{i+1} before reaching either A or B. In the last line we summed the expression realizing that $P(\lambda_i|\lambda_0) = \prod_{j=0}^{i-1} P(\lambda_{j+1}|\lambda_j)$.

Eq. 6.2 is a generic equation to calculate transition rate between two states. Eq. 6.2 reduces to Eq. 6.1, when the trajectories from the initial state have to pass through all the interfaces to reach the final state.

We note that a similar situation of stables states nested between interfaces occurs in multiple state TIS and single replica exchange TIS, which was treated in [12, 4, 7, 8].

6.2.2. EFFECTIVE AND INTRINSIC RATES FOR TWO BINDING SITES

We now apply the generic Eq. 6.2 to the specific case of an enzyme-substrate association-dissociation reaction, where the enzyme has one binding site while the substrate has two binding sites. We will first discuss the details of the path ensembles that correspond to each of the rate constants.

The intrinsic dissociation rate for the enzyme initially bound at patch A, $k_\mathrm{d}^A(\sigma)$, involves counting to the trajectories that start at A and reach σ without reaching the patch B. Similarly, the effective dissociation rate for the enzyme starting at patch A, k_off^A, takes into account all trajectories starting at A and going up to infinity, without visiting patch B. The effective association constant of an enzyme to bind to patch A, k_on^A, corresponds to trajectories that start at infinity and terminate at patch A without first visiting B, while the intrinsic association rate $k_\mathrm{a}^A(\sigma)$ corresponds to trajectories that start at the σ-interface and terminate at patch A without first visiting B. The intrinsic rate of hopping from patch A to B, $k_\mathrm{hop}^{AB}(\sigma)$, corresponds to trajectories that start in patch A and end in patch B, without first visiting the σ-interface, while the effective rate constant of hopping from patch A to patch B, $k_\mathrm{effHop}^{AB}(\sigma)$, corresponds to trajectories that start in patch

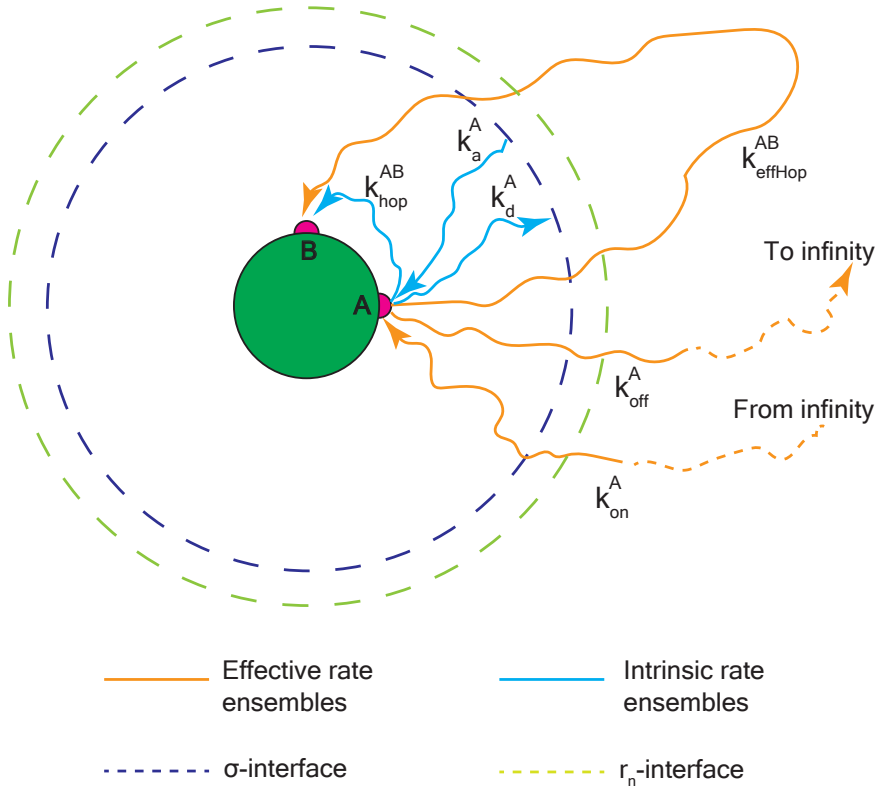


Figure 6.2: The path ensembles contributing to each rate constant. The substrate has two patches A and B and when the particle diffuses to infinity, it is in the unbound state. All rates are calculated with respect to the patch A, i.e. the rate of binding to patch A or the rate of unbinding from patch A. σ is the interface where the intrinsic rates are measured and r_n is an interface beyond σ.

A and end in patch B, without diffusing to infinity (however, they may cross the σ interface). We can similarly define the ensemble of trajectories that define the rate constants for associating to and dissociating from patch B. We emphasize that while the intrinsic rate constants depend on the definition of σ, the effective rate constants do not.

We will now show how from a single FFS simulation of a dissociation reaction from patch A, both the intrinsic and effective dissociation rates and the intrinsic and effective hopping and association rates can be computed. The intrinsic dissociation rate at the σ-interface is given by

$$k_d^A(\sigma) = \Phi_0 P(\sigma|r_0), \tag{6.3}$$

where, as before, Φ_0 is the flux across the first FFS interface r_0, and $P(\sigma|r_0)$ is the probability that an enzyme starting at this first interface reaches the σ-interface before binding to either patch, and can be computed using Eq. 6.1 . The effective dissociation constant

can be similarly expressed as,

$$k_{\rm off}^{A} = \Phi_0 P(\sigma|r_0) P(\infty|\sigma) = k_{\rm d}^{\rm A}(\sigma) P(\infty|\sigma), \tag{6.4}$$

where $P(\infty|\sigma)$ is the probability that a trajectory that has reached σ, escapes to infinity, i.e. dissociates, before it associates with either patch A or patch B. Below, we describe how $P(\infty|\sigma)$ can be obtained efficiently in an FFS simulation.

The intrinsic hopping rate given that the enzyme starts at patch A, is given by

$$k_{\rm hop}^{\rm AB} = \Phi_0 P(B|r_0), \tag{6.5}$$

where $P(B|r_0)$ is the probability that the enzyme, starting in patch A, associates with patch B before it arrives at σ.

The effective hopping rate, given that the enzyme starts bound to patch A, is,

$$k_{\rm effHop}^{AB} = \Phi_0 \left[P(B|r_0) + P(\sigma|r_0) P(B|\sigma) \right] \tag{6.6}$$

where $P(\sigma|r_0)$ is the probability of the enzyme starting at the first interface r_0, subsequently reaches the σ-interface before binding to either patch. $P(B|\sigma)$ is the probability that the enzyme, starting at σ, reaches B before it either escapes (i.e. dissociates) or reaches A; the path ensemble corresponding to this effective hopping rate, does include, however, trajectories that progress (significantly) beyond the σ interface.

To see how all the rate constants (intrinsic/effective association, dissociation and hopping rates) can be obtained from one single FFS simulation of a dissociation reaction, it is instructive to imagine that starting in the bound state A, we have generated N configurations at interface σ. These configurations are thus distributed over the σ surface according to the distribution as obtained from an FFS dissociation simulation starting from patch A.

Of the 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: $N = N_{\sigma\to \rm B} + N_{\sigma\to \rm A} + N_{\sigma\to\infty}$. In the limit of $N\to\infty$, we can define the probabilities $P(B|\sigma) = N_{\sigma\to B}/N$, $P(A|\sigma) = N_{\sigma\to A}/N$, $P(\infty|\sigma) = N_{\sigma\to\infty}/N$. For a finite number of sampled trajectories, the trajectory fractions become approximations of these probabilities. We can then write

$$\begin{aligned} P(B|\sigma) &= \frac{P(B|\sigma)}{P(A|\sigma)+P(B|\sigma)} \left(P(A|\sigma)+P(B|\sigma)\right) \\ &= \frac{P(B|\sigma)}{P(A|\sigma)+P(B|\sigma)} \left(1-P(\infty|\sigma)\right) \\ &= \alpha \left(1-P(\infty|\sigma)\right). \end{aligned} \tag{6.7}$$

where we have defined

$$\alpha \equiv \frac{P(B|\sigma)}{P(A|\sigma)+P(B|\sigma)} \tag{6.8}$$

and made use of the fact that $P(A|\sigma)+P(B|\sigma)+P(\infty|\sigma)) = 1$, and we have introduced α as a 'splitting probability' for trajectories from σ arriving at B versus A. To compute $P(\infty|\sigma)$ 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 mitigate this problem, we put, following our earlier work [1], an interface at a position $r_\mathrm{n} > \sigma$. As we will show below, this extra interface makes it possible to efficiently compute $P(\infty|\sigma)$. Moreover, the probability $P(B|\sigma)$ for trajectories that move from σ to B is then given by the sum of the probability $P_\mathrm{dir}(B|\sigma)$ of trajectories that directly go from σ to B without first visiting r_n, and the probability $P_{r_\mathrm{n}}(B|\sigma)$ for those that first visit r_n and then proceed to B,

$$\begin{aligned} P(B|\sigma) =& P_\mathrm{dir}(B|\sigma) + P_{r_\mathrm{n}}(B|\sigma) \\ =& \alpha_\mathrm{dir}\left(P_\mathrm{dir}(A|\sigma) + P_\mathrm{dir}(B|\sigma)\right) + \\ & \alpha_{r_\mathrm{n}}\left(P_{r_\mathrm{n}}(A|\sigma) + P_{r_\mathrm{n}}(B|\sigma)\right) \\ =& \alpha_\mathrm{dir}(1 - P(r_\mathrm{n}|\sigma)) + \alpha_{r_\mathrm{n}} P(r_\mathrm{n}|\sigma)(1 - P(\infty|r_\mathrm{n})). \end{aligned} \tag{6.9}$$

Here, α_dir and α_{r_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:

$$\alpha_\mathrm{dir} = \frac{P_\mathrm{dir}(B|\sigma)}{P_\mathrm{dir}(A|\sigma) + P_\mathrm{dir}(B|\sigma)} \tag{6.10}$$

$$\alpha_{r_\mathrm{n}} = \frac{P_{r_\mathrm{n}}(B|\sigma)}{P_{r_\mathrm{n}}(A|\sigma) + P_{r_\mathrm{n}}(B|\sigma)} \tag{6.11}$$

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

$$\begin{aligned} P(A|\sigma) = & (1 - \alpha_\mathrm{dir})(1 - P(r_\mathrm{n}|\sigma)) + \\ & + (1 - \alpha_{r_\mathrm{n}}) P(r_\mathrm{n}|\sigma)(1 - P(\infty|r_\mathrm{n})). \end{aligned} \tag{6.12}$$

As a sanity check, we can add Eq. 6.9 and Eq. 6.12 which gives

$$\begin{aligned} P(A|\sigma) + P(B|\sigma) &= 1 - P(r_\mathrm{n}|\sigma) + P(r_\mathrm{n}|\sigma)(1 - P(\infty|r_\mathrm{n}) \\ &= 1 - P(r_\mathrm{n}|\sigma) P(\infty|r_\mathrm{n}) \\ &= 1 - P(\infty|\sigma), \end{aligned} \tag{6.13}$$

which is indeed equal to the probability to observe trajectories that do not escape to infinity and hence bind to either A or B.

Combining Eqs. 6.8, 6.9 and 6.12 yields

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

The quantities $P(r_\mathrm{n}|\sigma)$, $P(\infty|\sigma)$ and α_dir can be directly obtained from the FFS simulations. Hence, to close the above equation and find $P(B|\sigma)$ (see Eq. 6.7), we need an expression for α_{r_n}. Since $P_{r_\mathrm{n}}(B|\sigma)$ is the product of the probability $P(r_\mathrm{n}|\sigma)$ of trajectories going from σ to r_n and the probability $P(B|r_\mathrm{n})$ of subsequently reaching B, the splitting

probability α_{r_n} in Eq. 6.11 is also given by

$$\begin{aligned}\alpha_{r_\mathrm{n}} &= \frac{P_{r_\mathrm{n}}(B|\sigma)}{P_{r_\mathrm{n}}(A|\sigma)+P_{r_\mathrm{n}}(B|\sigma)}\\ &= \frac{P(r_\mathrm{n}|\sigma)P(B|r_\mathrm{n})}{P(r_\mathrm{n}|\sigma)P(A|r_\mathrm{n})+P(r_\mathrm{n}|\sigma)P(B|r_\mathrm{n})}\\ &= \frac{P(B|r_\mathrm{n})}{P(A|r_\mathrm{n})+P(B|r_\mathrm{n})}\end{aligned} \tag{6.14}$$

We emphasize 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, α_{r_n} is likely to be unequal to α_dir, which would make it impossible to close Eq. 6.14. In contrast, if the distributions at the σ and the r_n interfaces are isotropic, then

$$\frac{P_{r_\mathrm{n}}(B|\sigma)}{P_{r_\mathrm{n}}(A|\sigma)+P_{r_\mathrm{n}}(B|\sigma)} \simeq \frac{P_\mathrm{dir}(B|\sigma)}{P_\mathrm{dir}(A|\sigma)+P_\mathrm{dir}(B|\sigma)} \tag{6.15}$$

and, thus

$$\alpha_{r_\mathrm{n}} = \alpha_\mathrm{dir}. \tag{6.16}$$

Inserting Eq. 6.16 in Eq. 6.14, we find

$$\alpha \equiv \frac{P(B|\sigma)}{P(A|\sigma)+P(B|\sigma)} = \alpha_\mathrm{dir}, \tag{6.17}$$

which reduces Eq. 6.7 to

$$P(B|\sigma) = \alpha_\mathrm{dir}(1-P(\infty|\sigma)). \tag{6.18}$$

Hence the effective hopping rate from Eq. 6.6 reduces to,

$$k^{AB}_\mathrm{effHop} = \Phi_0\left[P(B|r_0)+\alpha_\mathrm{dir}P(\sigma|r_0)(1-P(\infty|\sigma))\right]. \tag{6.19}$$

Eq. 6.4 shows that the effective dissociation rate k^A_off is given by an intrinsic dissociation rate k^A_d times an escape probability $P(\infty|\sigma)$. The escape probability, together with the diffusion-limited arrival rate $k_D(\sigma)$, makes it possible to define the intrinsic association rate [1]:

$$P(\infty|\sigma) = \frac{k_\mathrm{D}(\sigma)}{k_\mathrm{a}^{\mathrm{A}\vee\mathrm{B}}(\sigma)+k_\mathrm{D}(\sigma)}. \tag{6.20}$$

where $k_\mathrm{a}^{\mathrm{A}\vee\mathrm{B}}$ is the intrinsic rate at which a particle at σ binds either A or B. This equation, and hence the intrinsic association rate, can, in principle, also be defined for surfaces σ for which the distribution of configurations is not isotropic [1]; yet, the expression for the diffusion-limited arrival rate $k_D(\sigma)$ is then, in general, not known. In the case considered here, where the distribution of configurations at σ is isotropic (and σ is (significantly) beyond the cut-off of the potential), the diffusion-limited arrival rate is, however, simply given by $k_D(\sigma) = 4\pi\sigma D$, where D is the (sum of the) diffusion constants

of the particles. Rearranging the above equation yields the intrinsic rate of binding to either A or B,

$$k_{\mathrm{a}}^{\mathrm{A\vee B}}(\sigma) = k_{\mathrm{D}}(\sigma)\frac{1-P(\infty|\sigma)}{P(\infty|\sigma)}. \tag{6.21}$$

The effective association rate of binding to either A or B is given by the diffusion-limited arrival rate $k_D(\sigma)$ times the binding probability $(1-P(\infty|\sigma))$,

$$k_{\mathrm{on}}^{\mathrm{A\vee B}} = k_{\mathrm{D}}(\sigma)(1-P(\infty|\sigma)), \tag{6.22}$$

and the effective association rate of binding to patch A is

$$k_{\mathrm{on}}^{\mathrm{A}} = \alpha(1-P(\infty|\sigma))k_{\mathrm{D}}(\sigma). \tag{6.23}$$

and the intrinsic association rate of binding to patch A is given by,

$$k_{\mathrm{a}}^{\mathrm{A}}(\sigma) = \alpha k_{\mathrm{a}}^{\mathrm{A\vee B}}(\sigma) \tag{6.24}$$

Eqs. 6.3, 6.4, 6.5, 6.19, 6.24, and 6.23 yield the expressions for the intrinsic and effective dissociation, hopping and association rates. Only one point remains to be addressed, which is how $P(\infty|\sigma)$ can be obtained efficiently in an FFS simulation. To this end, we exploit that the effective rate of binding either patch is independent on the location of the effective cross section [1]:

$$k_{\mathrm{on}}^{\mathrm{A\vee B}}(\sigma) = k_{\mathrm{on}}^{\mathrm{A\vee B}}(r_{\mathrm{n}}), \tag{6.25}$$

which means, using Eq. 6.22, that

$$k_{\mathrm{D}}(\sigma)(1-P(\infty|\sigma)) = k_{\mathrm{D}}(r_{\mathrm{n}})(1-P(\infty|r_{\mathrm{n}})). \tag{6.26}$$

$P(\infty|\sigma)$ can be factorized as

$$P(\infty|\sigma) = P(\infty|r_{\mathrm{n}})P(r_{\mathrm{n}}|\sigma) \tag{6.27}$$

Solving Eq. 6.26 and Eq. 6.27 yields

$$P(\infty|\sigma) = \frac{P(r_{\mathrm{n}}|\sigma)(1-\Omega)}{1-\Omega P(r_{\mathrm{n}}|\sigma)}, \tag{6.28}$$

where $\Omega \equiv k_{\mathrm{D}}(\sigma)/k_{\mathrm{D}}(r_{\mathrm{n}})$. 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 $k_{\mathrm{D}}(\sigma) = 4\pi\sigma D$ to evaluate all the rate constants.

The above equations hold for situation in which both A and B are very long lived states, so that the association rate k_{AU} for binding to A is dominated by the paths that directly proceed from the unbound state to path A, and the paths that visit B do not contribute significantly to k_{AU}. We argue that also in the (mean-field) modelling of biochemical networks this is the most natural and useful definition of the association rate. However, in some cases, one may wish to treat B as an intermediate state for the binding to and the unbinding from A.

6.2.3. PARTICLE MODEL AND INTERACTION POTENTIAL

All particles are spherical with an isotropic centre of mass interaction, dressed with one or more sticky spots on their surface called 'patches', which allow for highly directional, anisotropic interactions. We use two species of such particles in our simulations, a substrate particle which has two patches and an enzyme particle that has one patch.

The enzyme-substrate pair, in our model, experiences a strong attractive potential, $U_{\mathrm{s}}(r)$, over a narrow band of orientations (see Fig. 6.3). 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 a repulsive potential, $U_{\mathrm{rep}}(R)$, which is a function of the center-of-mass distance, R. In addition, particles experience a weak, isotropic, non-specific attraction, $U_{\mathrm{ns}}(R)$. The total patch potential reads:

$$U_{\mathrm{an}}(R,r) = \sum_{\mathrm{i}=1}^{\mathrm{n}} U_{\mathrm{s}}(r_{\mathrm{i}}) + U_{\mathrm{rep}}(R) + U_{\mathrm{ns}}(R), \tag{6.29}$$

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

$$U_i(x) = \begin{cases} \epsilon_i \left(1 - a_i \left(\frac{x}{\sigma_{\mathrm{an}}}\right)^2\right) & \text{if } x < x_i^\star, \\ \epsilon_i b_i \left(\frac{x_i^c}{\sigma_{\mathrm{an}}} - \frac{x}{\sigma_{\mathrm{an}}}\right)^2 & \text{if } x_i^\star < x < x_i^c, \\ 0 & \text{otherwise,} \end{cases} \tag{6.30}$$

with $i = \{\mathrm{s}, \mathrm{rep}, \mathrm{ns}\}$ respectively. The overall strength ϵ_i, the length scale σ_{an}=5nm, 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 set the following parameters: $\epsilon_{\mathrm{s}} = 20 k_{\mathrm{B}} T$, $a_{\mathrm{s}} = 20$ and $r_{\mathrm{att}}^* = 0.1\sigma_{\mathrm{an}}$, implying $b_{\mathrm{s}} = 5$ and $r_{\mathrm{s}}^c = 0.5\sigma_{\mathrm{an}}$; $\epsilon_{\mathrm{rep}} = 100 k_{\mathrm{BT}}$, $a_{\mathrm{rep}} = 1$ and $R_{\mathrm{rep}}^* = 0.85\sigma_{\mathrm{an}}$, implying $b_{\mathrm{rep}} = 2.6036$ and $R_{\mathrm{rep}}^c = 1.1764\sigma_{\mathrm{an}}$; and $a_{\mathrm{ns}} = 1$ and $R_{\mathrm{ns}}^* = 0.85\sigma_{\mathrm{an}}$, implying $b_{\mathrm{ns}} = 2.6036$ and $R_{\mathrm{ns}}^c = 1.1764\sigma_{\mathrm{an}}$. ϵ_{ns} is varied from $2 k_{\mathrm{B}} T$ to $20 k_{\mathrm{B}} T$ with steps of $2 k_{\mathrm{B}} T$. Fig. 5.4b shows the total potential as a function of R, when the patches are aligned ($r = R - \sigma_{\mathrm{an}}$) and misaligned ($r = R + \sigma_{\mathrm{an}}$). 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, $U_{\mathrm{s}} = 0$ and the particles only experience the weak U_{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.

Particles of the same species, i.e enzyme-enzyme and substrate-substrate only have a WCA repulsion based on the centre of mass distance, R.

6.3. RESULTS

In this section, we evaluate the transition rates using the expressions derived in the section 6.2.2 for the patchy particle model system described in the 6.2.3. First, we determine all rate constants for the case with the enzyme initially bound to a one of the patches

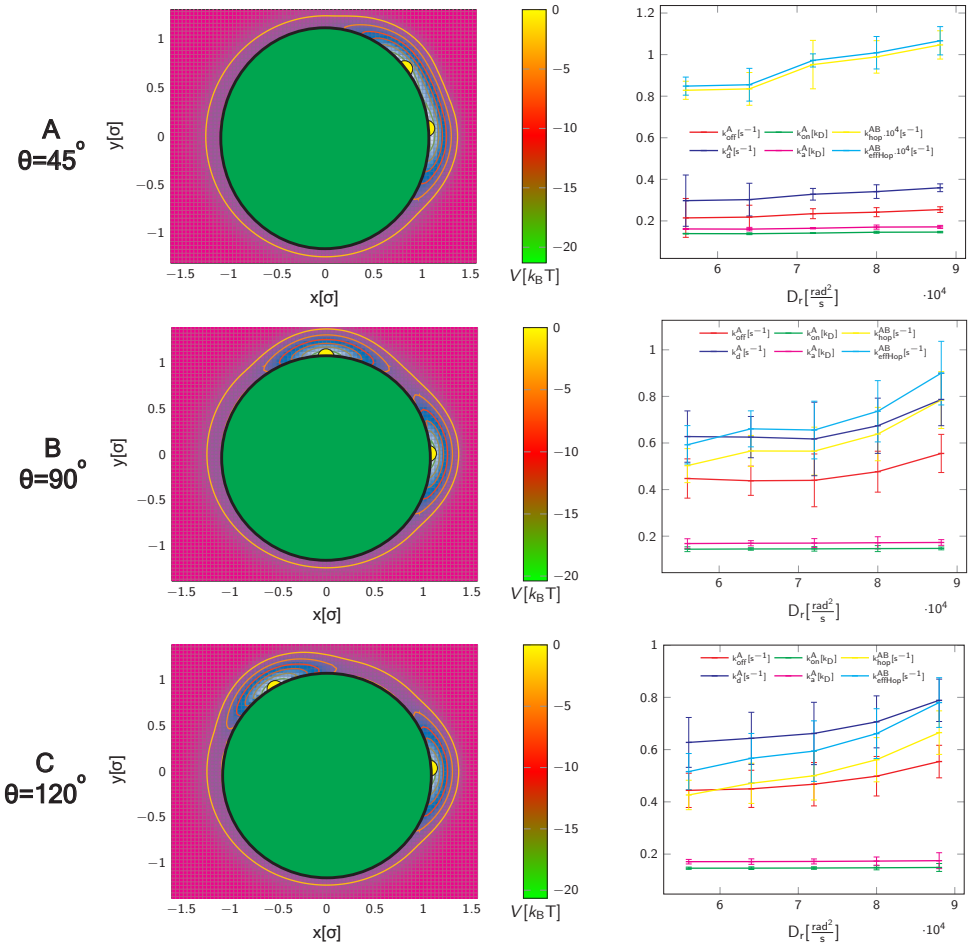


Figure 6.3: Potential energy landscape (left column) of the enzyme-substrate system as a function of the spacial position of the pair, and rate constants evaluated with the derived expressions(right column). The energy landscape is obtained by moving the enzyme around the substrate, with the patch of the enzyme pointed to the centre of the substrate. The rows A, B, C correspond to a patch spacing θ (angle between the two patch vectors) of $45°, 90°$ and $120°$ respectively. For all three angles all the rates increase as a function of the orientational diffusion constant D_r. However the hopping rates show the strongest dependence, while the association rates the least. The hopping rates are largest when $\theta = 45°$ and decreases as the spacing increases. The dissociation rates are smallest when $\theta = 45°$, increases when $\theta = 90°$. Further, increase in the spacing between the patches has no effect on the dissociation rate. The association rates are not influenced by the patch spacing.

of the substrate. Employing FFS on the dissociation reaction we determine all six rate constants as a function of the rotational diffusion constant D_r for several patch angu-

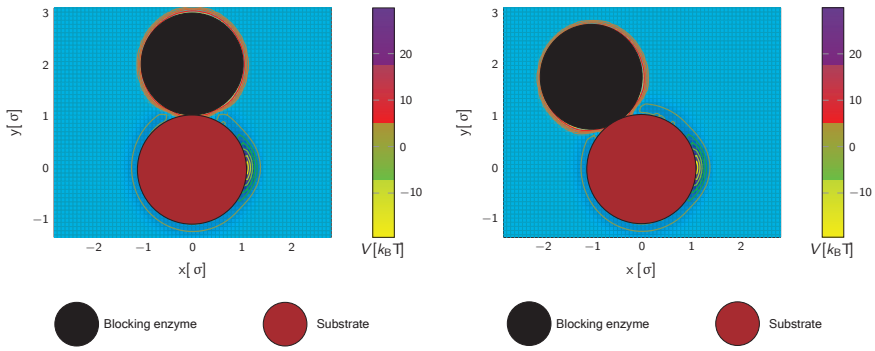


Figure 6.4: Potential energy landscapes when one of the patches is blocked by an enzyme when $\theta = 90°$(left) and $\theta = 120°$(right). The second enzyme particle(not shown in the figure), feels a strong attraction around the free patch, a weak non-specific attraction around the substrate and a repulsive force around the blocking enzyme.

lar spacings. Next, we compare the computed association and dissociation rates for this case to two other scenarios: (i) the other substrate patch is blocked by a static/inert (second) enzyme (ii) the substrate has only a single patch, which is identical to the other patch being blocked by an infinitesimally small enzyme.

All simulations are performed using a Brownian dynamics integrator [13]. The system specific parameters of the simulation are as follows: The particle diameter is $\sigma =$ 5nm, the time step $\delta t = 0.1$ns, the mass of the particle is $m = 50$kDa, the mass moment of inertia $M = \frac{8}{15} m\sigma^2$, the translational and rotational friction coefficients are $\gamma = \frac{k_B T}{D_t m}$ the $\Gamma = \frac{k_B T}{D_r M}$ respectively, where D_t and D_r translational and rotational diffusion constants, $k_B = 1.38 \times 10^{-23}$JK^{-1} is the Boltzmann constant and $T = 300$K is the temperature of the system.

6.3.1. EFFECT OF THE ROTATIONAL DIFFUSION CONSTANT AND PATCH SPACING ON THE RATE CONSTANTS

From the FFS simulation of the dissociation reaction of an enzyme initially bound to a patch we compute the association, dissociation and the hopping rate constants as a function of the rotational diffusion constant for patch opening angles $45°, 90°$ and $120°$. Fig. 6.3 shows the potential energy landscapes and the rate constants resulting from Eqs. Eqs. 6.3, 6.4, 6.5, 6.19, 6.24, and 6.23.

For all path opening angles the rate constants only weakly increase as a function of the orientational diffusion constant D_r. The hopping rate constant has the strongest dependence on D_r, whereas the association rate constant has the weakest. At higher D_r the particles rotate faster, allowing the particle to leave the potential well more easily, since even a small misalignment of the patches causes unbinding. Once unbound, the particle either dissociates or hops to the other patch. Hence, the hopping and dissociation rate constants sensitively depend on D_r. The association rate constant, on the other

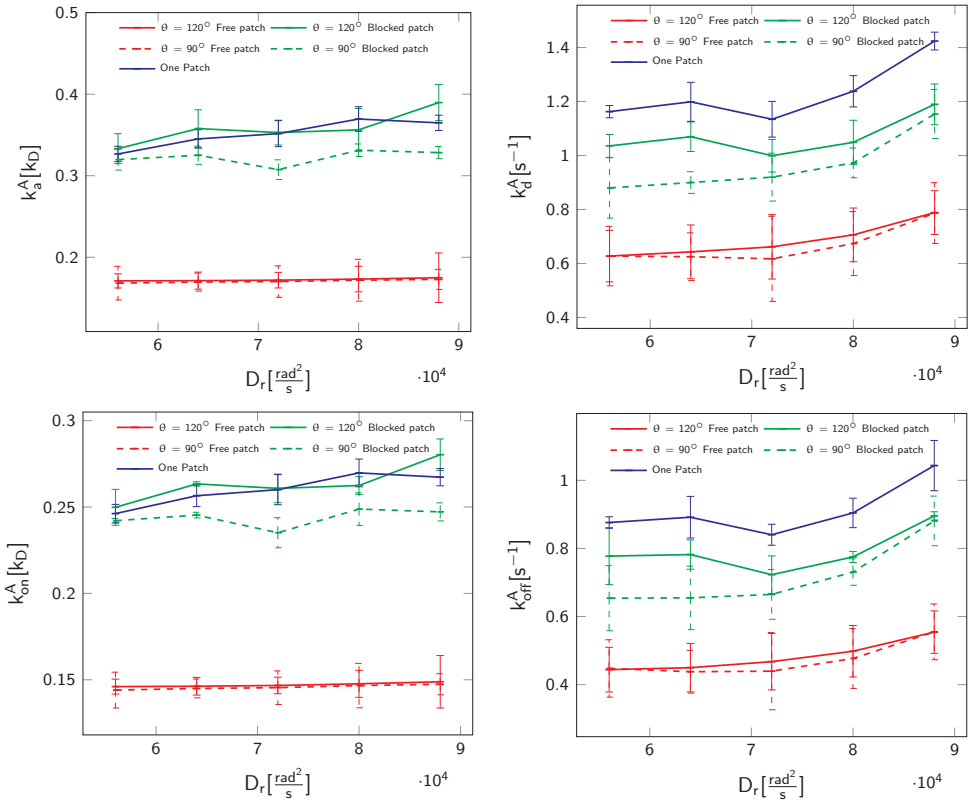


Figure 6.5: Clockwise from top left: The intrinsic association rate, k_a, intrinsic dissociation rate, k_d, effective dissociation rate, k_{off} and effective association rate, k_{on}, plotted as a function of the rotational diffusion constant, D_r, for three cases: (i) substrate with two patches and both patches are free (Free patch), (ii)substrate with two patches and one of the patch is blocked, (iii) substrate with one patch. The association rates are largest when the substrate has one patch and when one patch is blocked with $\theta = 120°$. The association rates for the blocked patch when $\theta = 90°$ is lower than the above value and is least when the substrate has two patches with no blocking. The dissociation rate on the other hand is largest when the substrate has a single patch and least when both patches are free. When one patch is blocked the dissociation rates lie in between the above values.

hand increases only marginally with D_r, as the rotational diffusion does not limit the rate constant of association for these values of D_r.

Fig. 6.3A shows that for the $\theta = 45°$ case, the two patches partly overlap, enhancing the probability for hopping rather than dissociation. Hence, for $\theta = 45°$ the dissociation rate constant is lower and the hopping rate constant is orders of magnitude larger when compared to the dissociation and the hopping rate constants when $\theta = 90°$ or $\theta = 120°$. For the higher patch angles the two patches truly separate, so that an enzyme unbinding from one patch rarely hops to the other patch. Whenthe patch angle is increased from

$\theta = 45^\circ$ to $\theta = 90^\circ$, the dissociation rate constant initially increases but levels off when the patches are positioned at $\theta = 120^\circ$. Association rate constants are hardly dependent on the angular distance between the patches.

6.3.2. EFFECT OF BLOCKING ON THE RATE CONSTANTS

As the substrate has more than one binding site (patch), multiple enzymes can bind simultaneously. The presence of another bound enzyme might affect the association association and dissociation rate constants of an enzyme to and from the free patches. Restricting ourselves to a substrate with two patches we consider first the case where one of the two patches is blocked by an identical enzyme. We assume this enzyme to remain bound and treat it as being static and inert. In this case, the hopping rate constants are zero since the enzyme cannot hop to the other patch. We compute the effective and intrinsic association/dissociation rate constants for patch opening angles $\theta = 90^\circ$ and $\theta = 120^\circ$ We also compare the computed rate constants to those obtained when there is no blocking and the case where the substrate has only one patch.

Fig. 6.4 shows the energy landscapes for $\theta = 90^\circ$ and $\theta = 120^\circ$, while Fig. 6.5 shows the intrinsic and effective association and dissociation rate constants for all cases. From the energy landscape we see that the mobile enzyme(not in figure) feels a strong specific attractive force around the free patch, a weak non-specific attractive force around the substrate and a repulsive force near the blocking enzyme. Fig. 6.5 shows the association rate constants k_a^A and k_{on}^A are highest for the one patch case and for blocked patch case with $\theta = 120^\circ$. For the blocked patch case with $\theta = 90^\circ$ the association rate constants are slightly smaller because the blocking enzyme is closer to the free patch and causes steric hindrance, reducing the association rate constant. When the patches are further apart, this effect is reduced, corresponding to an increased association rate constant. The association rate constants are lowest for the unblocked two patch case. Since the association is specific to patch A, we do not count trajectories that bind to patch B, which acts as a trap, effectively reducing the association rate constant for binding to patch A.

The dissociation rate constants are largest for the one patch situation, because once the enzyme leaves the potential well, it is only held by the weak non-specific attraction, enhancing the chance to escape. The dissociation rate constants are lowest for the unblock two-patch case, since an enzyme leaving a potential well still can hop to the other patch and rebind to the substrate, thus reducing the dissociation rate constants. The dissociation rate constants for the blocked-patch scenario is in between the two other cases, as then an enzyme leaving a potential well is repelled by the blocking enzyme and rebinds to the patch where it started from. This effect is larger for small distances between the patches, leading to smaller dissociation rate constants for $\theta = 90^\circ$ compared to the dissociation rate constant for $\theta = 120^\circ$.

6.4. CONCLUSION

In this work, we derived a generic expression to evaluate the dissociation rate constant using FFS, for cases where two states are not necessarily separated by all interfaces. Next, we derive microscopic expressions for intrinsic and effective association, dissociation and hopping rate constants, to be used in conjunction with a single rare-event simula-

tion of the dissociation reaction.

Because in signalling networks, the rebinding of the enzyme to the substrate can significantly change the response of the system, it is interesting to study the rate constants of binding, unbinding and hopping(rebinding) of an enzyme to the substrate. In our model, we restrict the number of binding sites on the substrate to two and the enzyme has one binding site. For this model we calculate 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 rate constants are mostly independent of how fast the particles rotate, while the dissociation and hopping rate constants are more strongly correlated with D_r. When the patches are close to each other, the enzyme hops(rebinds) to the other substrate patch, instead of diffusing away. This hopping rate constant depends on the patch distance very strongly. In contrast, the association and dissociation rate constants do not change significantly with patch separation. Finally, we studied the effect of a blocking enzyme on the associating/dissociation/hopping rate constants. In the presence of an (inert) blocking enzyme, the association rate constants decrease and the dissociation rate constants increase when compared to the unblocked case. 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 multiscale simulation[5, 6], where, by using these rate constants, explicit simulations of the association reactions can be avoided, which dramatically speeds up the simulations.

6.5. ACKNOWLEDGEMENTS

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.

BIBLIOGRAPHY

[1] Vijaykumar A, Bolhuis PG, ten Wolde PR (2016) The intrinsic rate constants in diffusion-influenced reactions. *Faraday Discussions* 195:421–441.

[2] Pan AC, Borhani DW, Dror RO, Shaw DE (2013) Molecular determinants of drug–receptor binding kinetics. *Drug Discovery Today* 18:667 – 673.

[3] Takahashi K, Tănase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America* 107:2473–8.

[4] Newton AC, Groenewold J, Kegel WK, Bolhuis PG (2015) Rotational diffusion affects the dynamical self-assembly pathways of pa tchy particles. *Proc. Nat. Acad. Sci.* 112:15308–15313.

[5] Vijaykumar A, Bolhuis PG, ten Wolde PR (2015) Combining molecular dynamics

with mesoscopic Green's function reaction dynamics simulations. *The Journal of Chemical Physics* 143:214102.

[6] Vijaykumar A, Ouldridge TE, ten Wolde PR, Bolhuis PG (2017) Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics. *The Journal of Chemical Physics* 146:114106.

[7] Newton AC, Groenewold J, Kegel WK, Bolhuis PG (2017) *J. Chem. Phys* p in press.

[8] Newton AC, Kools R, Swenson DW, Bolhuis PG (2017) *J. Chem. Phys.* p submitted.

[9] Allen RJ, Warren PB, ten Wolde PR (2005) Sampling Rare Switching Events in Biochemical Networks. *Physical Review Letters* 94:018104.

[10] Allen RJ, Valeriani C, Rein ten Wolde P (2009) Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter* 21:463102.

[11] van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. *The Journal of Chemical Physics* 118:7762.

[12] Du W, Bolhuis PG (2014) Sampling the equilibrium kinetic network of trp-cage in explicit solven t. *J. Chem. Phys.* 140:195102.

[13] Davidchack R, Ouldridge T, Tretyakov M (2015) New langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* 142.

7

MULTI-SCALE SIMULATIONS OF MULTI-SITE PHOSPHORYLATION IN A MAPK SIGNALLING PATHWAY

In many reaction-diffusion processes, the spatial distribution of the reactants and the stochastic character of their interactions are crucial for the macroscopic behaviour. A prime example is multi-site protein modification, as in the MAPK system. To efficiently simulate such systems, it is necessary to use a particle based modelling approach. The obvious choice is to use a particle based simulation technique based on Brownian Dynamics. Brownian Dynamics can capture the dynamics at the microscopic scales naturally. However, when the concentration of the system is low, the technique becomes very inefficient since it spends most of the time bringing the particles close to each other. The recently developed mesoscopic Green's Function Reaction Dynamics (GFRD) method enables efficient simulation at the particle level provided the microscopic dynamics can be integrated out. Yet, many processes exhibit non-trivial microscopic dynamics that can qualitatively change the macroscopic behavior, calling for an atomistic, microscopic description. In our previous work, we present a multi-scale scheme Molecular Dynamics - Green's Functions Reaction Dynamics (MD-GFRD), that combines the computational power of GFRD and the microscopic detail of BD. MD-GFRD simulates the particles with GFRD when the particles are far apart and switches to BD to capture the microscopic dynamics when the particles come close to each other. In this paper, we use MD-GFRD to simulate the MAPK signalling pathway and determine the response of the network. Our study will enable a systematic study of the response of the network as a function of the distance between the binding sites, the rotational diffusion constant and the ratio of the substrate to the enzyme concentration.

7.1. INTRODUCTION

It is now clear that in many reaction-diffusion systems, events at the microscopic scale can qualitatively change the behaviour at the macroscopic scale [1]. A prime example is given by the phenomenon of enzyme-substrate rebindings in cellular systems that employ multi-site protein modification. Multi-site protein modification is a very common motif in cellular signaling. Examples are the Kai system [2], the CDK inhibitor Sic1 [3], the NFAT system [4], and the CAMKII system [5]. But arguably the most studied and best characterized example is provided by the MAPK signaling cascade, which is involved in cell differentiation, cell proliferation, and apoptosis [6]. Recently, Takahashi and coworkers showed that in these systems enzyme-substrate rebinding at the molecular scale can qualitatively the response of the system at the cellular scale [1]. However, they studied a rather coarse-grained model of the MAPK system, in which the particles are modeled as ideal spherical objects with isotropic interactions. It thus remains an open question how strong the effect of enzyme-substrate rebinding is in a more detailed model that takes into account the orientational dependence of the interaction potential and the effect of the orientational diffusion of the proteins.

A cascade of the MAPK system is made up of three layers. In each layer, an enzyme (a kinase) activates a substrate (also a kinase) of the next layer. A substrate is fully activated when both its sites are phosphorylated. A kinase phosphorylates and a phosphatase dephosphorylates a site of the substrate. The response of the system changes based on whether these enzymes modify the substrate processively or distributively [7, 8, 9]. A distributive mechanism is one where the enzyme, after modifying the first site, has to unbind from the substrate, diffuse into the bulk and undergo another independent encounter with the substrate, to bind to the next site in order to modify it. On the other hand, in a processive mechanism, the enzyme, after modifying a site, remains bound to the substrate before it modifies the next site, thus requiring only a single encounter with the substrate. Therefore, models based on mean-field chemical rate equations predict that an ultra-sensitive response is generated by a distributive mechanism. This is because the rate of full activation depends quadratically on the concentration of the upstream kinase [7, 8, 9]. In contrast, ultra sensitivity is lost for a processive mechanism. Moreover, when the enzymes are present in limiting amounts, then, according to these mean-field models, a distributive mechanism can generate bistability, while a processive scheme cannot [10].

However, Takahashi and coworkers showed that spatio-temporal fluctuations at the molecular scale can drastically change the predictions of the mean-field models [1]. Particle-based simulations using the GFRD algorithm revealed that enzyme-substrate rebinding can effectively turn a distributive mechanism into a processive one. This is because the probability that an enzyme molecule which has just dissociated from a substrate molecule, rebinds that molecule instead of diffusing away, does not depend on the concentration. Indeed, these rebinding events are not described by conventional chemical kinetics. As a result, the response of the system does not depend on whether the enzyme and substrate molecule remain physically bound to each other in between the modifications of the two sites, but rather on whether both sites are modified by one and the same enzyme molecule—an enzyme molecule that dissociates and then rebinds gives the same response as one that remains bound. The prediction of Takahashi *et al.*

was confirmed experimentally by Matsuda and coworkers [11]. A theoretical framework for describing the behavior of the system in the presence of rebindings was provided by Gopich and Szabo [12]. The implication of pseudo-processivity for kinetic proofreading was studied in Ref. [13].

The model of Takahashi *et al.* assumed that the reactants are highly idealized, spherical objects, interacting via isotropic interactions [1]. Here, we address the question of how the response depends on the distance between the interaction sites on the substrate and on the orientational diffusion constants of the particles. To this end, we apply our recently developed Molecular Dynamics - Green's Function Reaction Dynamics (MD-GFRD) technique [14, 15]. MD-GFRD is an exact scheme that makes it possible to simulate systems with complex interaction potentials. In the multi-scale MD-GFRD scheme, we use BD to simulate the reactants that are within a reaction distance, and we use GFRD to propagate the particles that are beyond this distance. MD-GFRD thus combines the power of GFRD to make large jumps in time and space when the particles are far away from each other with the ability of BD to resolve the microscopic dynamics when they are close. The regions where the particles are close together and simulated with high microscopic resolution, and the parts of space where they are far apart and simulated with lower mesoscopic resolution, are defined adaptively on the fly. This makes MD-GFRD, under biologically relevant conditions, orders of magnitude more efficient than brute-force BD with the same microscopic resolution [14, 15]. While we use Brownian Dynamics for simulating the particles at close distances, the principal idea is generic. In the scheme presented here, GFRD can also be combined with other schemes such as atomistic Molecular Dynamics (MD) or Markov State Modelling (MSM).

In this chapter, we present a proof-of-principle demonstration that MD-GFRD can simulate relatively complex biochemical networks such as the MAPK system. An extensive analysis of the behaviour or of the system as a function of the orientational diffusion constant and the distance between the patches is on-going work.

7.2. SYSTEM

While the MD-GFRD scheme is generic, it will be illuminating to describe the various steps of the algorithm in the context of the system that we aim to simulate here. We therefore first describe the system.

The system that we will study is one layer of the MAPK pathway, consisting of one dual modification cycle. The system thus consists of a kinase with one binding patch, a phosphatase with one binding patch, and a substrate with two patches, corresponding to the two phosphorylation sites. These molecules move with a given translational diffusion constant D_t and with a given orientational diffusion constant D_r. Phosphorylation and dephosphorylation proceeds via Michaelis-Menten kinetics and according to

a distributive mechanism. The reactions of the system are:

$$S + K \underset{k_d}{\rightleftharpoons} SK \overset{k_p}{\rightarrow} S_pK \rightarrow K + S_p \tag{7.1}$$

$$S_p + K \underset{k_d}{\rightleftharpoons} S_pK \overset{k_p}{\rightarrow} S_{pp}K \rightarrow K + S_{pp} \tag{7.2}$$

$$SK + K \underset{k_d}{\rightleftharpoons} KSK \overset{k_p}{\rightarrow} KS_pK \rightarrow K + S_pK \tag{7.3}$$

$$S_pP + K \underset{k_d}{\rightleftharpoons} KS_pP \overset{k_p}{\rightarrow} KS_{pp}P \rightarrow S_{pp}P + K \tag{7.4}$$

$$S_{pp} + P \underset{k_d}{\rightleftharpoons} S_{pp}P \overset{k_{dp}}{\rightarrow} S_pP \rightarrow P + S_p \tag{7.5}$$

$$S_p + P \underset{k_d}{\rightleftharpoons} S_pP \overset{k_{dp}}{\rightarrow} SP \rightarrow P + S \tag{7.6}$$

$$S_{pp}P + P \underset{k_d}{\rightleftharpoons} PS_{pp}P \overset{k_{dp}}{\rightarrow} PS_pP \rightarrow P + S_pP \tag{7.7}$$

$$S_pK + P \underset{k_d}{\rightleftharpoons} KS_pP \overset{k_{dp}}{\rightarrow} KSP \rightarrow SK + P \tag{7.8}$$

The top four reactions describe the phosphorylation of the substrate by the kinase, while the bottom four describe the dephosphorylation of the substrate by the phosphatase. For this study we assume that the system is symmetric: (1) the rate constants for phosphorylation, k_p, equal those of dephosphorylation, k_{dp}; (2) the two phosphorylation sites on the substrate are identical. Even though the sites are identical, it will be important to also include an event that is not listed in the reactions above, namely a hopping event, in which an enzyme dissociates from one patch and then rebinds to the other patch before it reaches a predefined distance from the substrate. This is because the dynamics of enzyme-substrate binding and unbinding can generate spatio-temporal correlations that will also depend on the (history of the) hopping events.

The phosphorylation reactions and dephosphorylation reactions are modeled as simple uni-molecular reactions, characterized with *given* rate constants k_p and k_{dp} (which we here take to be equal): these rates are part of the model specification, and hence input parameters for the simulations. The association, dissociation and hopping rates are determined by the interaction potential, and the translational and orientational diffusion constants; the latter three are part of the model definition ("input"), but the rates are not. The principal idea of MD-GFRD is that association, dissociation, and hopping are simulated via brute-force BD, meaning that these rates are naturally captured in the simulations. However, since the interaction between the enzyme and substrate is strong, dissociation and hopping are rare events. While we could thus simulate dissociation and hopping explicitly on the fly, most CPU time would be wasted on propagating the enzyme and the substrate particles while they would simply rattle around each other in the potential well. To circumvent this problem, we will pre-compute the dissociation and hopping rate by simulating two particles using a rare-event technique as described below, and then use the pre-computed rates in the final MD-GFRD simulations. In the MD-GFRD simulations, these rate constants have thus become "input", and for this rea-

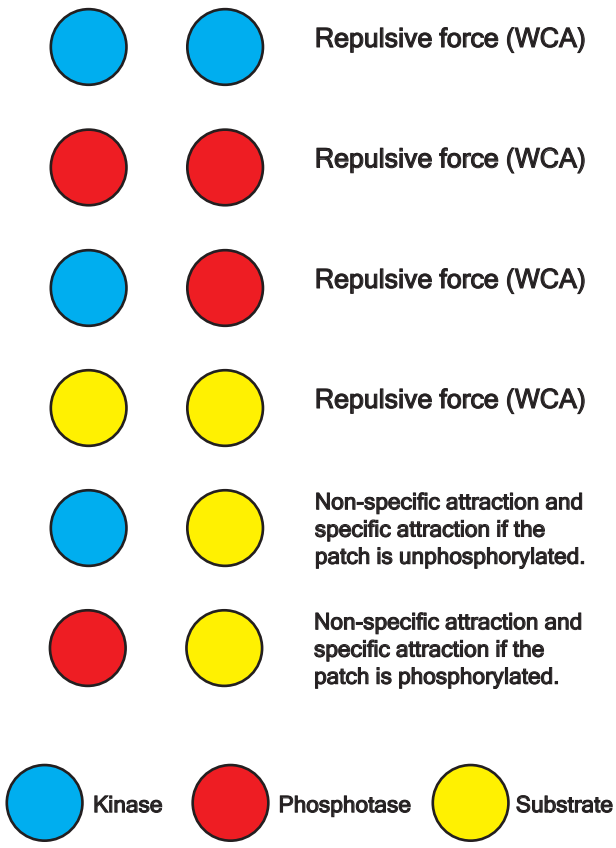


Figure 7.1: The interaction potentials between different species in the simulation. A pair of enzymes (kinases or phosphatases or one of each kind) and a pair of substrates experience a WCA repulsive force. A kinase is attracted to a substrate non-specifically and can bind specifically to a patch, only if the patch is unphosphorylated. Similarly a phosphatase is attracted to a substrate non-specifically and can bind specifically to a patch, only if the patch is phosphorylated.

son we have added a subscript k_d to the reactions above. We iterate, however, that these rates follow from the microscopic dynamics on the interaction potential, which is part of the model definition.

The interaction potential between an enzyme and a substrate molecule depends on the modification state of the phosphorylation sites of the substrate molecule. A kinase can have a specific—a strong and highly directional—attraction with only an *un* phosphorylated patch of the substrate and a phosphatase can have a specific attraction with only a phosphorylated patch of the substrate. In addition to these specific interactions, both the kinase and the phosphatase also interact non-specifically with the substrate; this non-specific interaction is much weaker and independent of both the orientation of the particles and the modification state of the substrate. All other interactions, i.e.

substrate-substrate and enzyme-enzyme are purely repulsive, modelled as a WCA repulsion. The interactions between different species is shown in Fig. 7.1. For a detailed description of the interaction potentials and particle models, we refer to Sec 3.2.2. and Sec 3.2.7. In summary, and for ease of description below, the complexes that are possible are thus: kinase-substrate and phosphatase-substrate complex, both described as an enzyme-substrate (ES) complex in the sections below; kinase-substrate-kinase, phosphatase-substrate-phosphatase, and kinase-substrate-phosphatase complex, described as enzyme-substrate-enzyme (ESE) complex — since the two patches are identical, there is only one unique kinase-substrate-phosphatase complex.

7.3. METHODS

7.3.1. MOLECULAR DYNAMICS – GREEN'S FUNCTIONS REACTION DYNAMICS (MD-GFRD)

MD-GFRD is an exact multiscale scheme that can straightforwardly be applied to any reaction-diffusion system with patchy particles. When particles are far apart, each particle is put into a non-overlapping spherical GFRD domain. Since the GFRD domains do not overlap, the stochastic processes in these domains are independent of one another. For each of the domains, the reaction-diffusion problem can be solved analytically using Green's functions. This yields for each domain an event *type*, which is either a reaction or an escape of a particle from the domain, and an event *time*, which is when this event will happen. These events are put in an event list, which is updated in chronological order. GFRD is thus an event-driven, asynchronous algorithm. A detailed description of GFRD is given in 2.2.2.

When the particles come close to each other, MD-GFRD switches to a microscopic technique like conventional BD. The GFRD domains around the particles are burst and the particles are collectively propagated with small time steps. The particles interact with each other via an interaction potential. This naturally allows for anisotropic particles and a description for conformational and orientational dynamics. When an enzyme molecule and a substrate molecule associate, the dimer is replaced by a single particle, which can be propagated with GFRD. After association, the enzyme may dissociate, hop from one site to the other, or phosphorylate/dephosphorylate the site. The dissociation/hopping reaction is not done explicitly in MD-GFRD. Instead, a pre-simulation is performed with just two particles using BD. To efficiently sample the dissociation and hopping reactions, which are rare events, Forward Flux Sampling (FFS) is used. The FFS simulation gives the dissociation rate k_d and hopping rates k_h, which describe exponential relaxation of these unimolecular reactions. Knowing the rate constants for the dissociation and hopping reactions, the next time for each of these reactions, τ_α with $\alpha = \mathrm{d},\mathrm{h}$, can be sampled from

$$\tau_\alpha = -k_\alpha \ln(\mathscr{R}), \tag{7.9}$$

where $\mathscr{R} \in [0,1]$ is a uniformly distributed random number. This step is discussed in detail in Sec. 7.3.3. For a detailed description of the coupling between the two schemes, we refer Sec. 3.2.5.

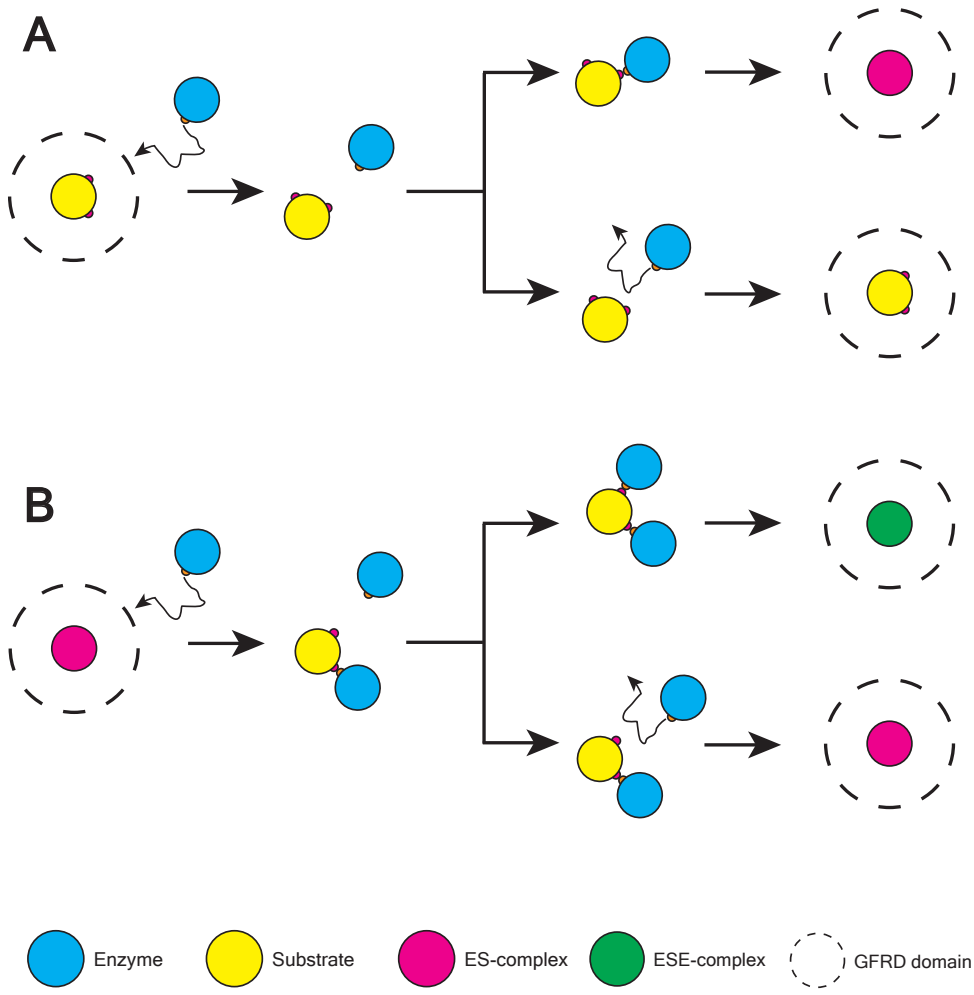


Figure 7.2: Handling association reactions in MD-GFRD. A. When an enzyme, comes closer than $\mathscr{D}_{\text{min}}$ to the surface of a GFRD domain that has a substrate, the domain is burst. The substrate and the enzyme is simulated explicitly with BD. Now one of two things may happen: (i) the enzyme can bind to the substrate (ii) the enzyme diffuses away from the substrate. If the enzyme binds to one of the patches of the substrate, then the enzyme-substrate dimer is replaced by a single particle, the enzyme-substrate (ES) complex. Now a domain can be constructed on the ES-complex, if there is space and is propagated with GFRD. On the other hand, if the enzyme diffuses away from the substrate, such that the distance between them is larger than $2\mathscr{D}_{\text{min}}$, then a domain is built around the substrate, and it is propagated with GFRD. B. When an enzyme, comes closer than $\mathscr{D}_{\text{min}}$ to the surface of a GFRD domain that has an ES-complex, the domain is burst. The EA-complex is split into its constituents i.e. the enzyme-substrate dimer which along with the enzyme are simulated explicitly with BD. At this point one of two things may happen: (i)the enzyme can bind to the free patch of the substrate (ii) the enzyme diffuses away from the enzyme-substrate dimer. If the enzyme binds to the free patch of the substrate, then the enzyme-substrate-enzyme trimer is replaced by a single particle, the enzyme-substrate-enzyme (ESE) complex. Now a domain can be constructed on the ESE-complex, if there is space and is propagated with GFRD. However, if the enzyme diffuses away from the enzyme-substrate dimer, such that the distance between them is larger than $2\mathscr{D}_{\text{min}}$, then the enzyme-substrate dimer is replaced by a single particle, the ES-complex. A domain is now built around the substrate, and it is propagated with GFRD.

7.3.2. HANDLING ASSOCIATION REACTIONS

When the distance between two particles is smaller than $2\mathscr{D}_{\text{min}}$, where $\mathscr{D}_{\text{min}}$ is the minimum domain radius, the particles are explicitly simulated with brute force rotational BD. If an enzyme and a substrate molecule happen to associate as shown in Fig. 7.2.A, the resulting enzyme-substrate dimer is replaced by a single particle, the enzyme-substrate (ES) complex in Fig. 7.2). This ES-complex is now placed in a GFRD domain, if there is space to build a domain of radius $2\mathscr{D}_{\text{min}}$. It could happen that another particle approaches this ES-complex in the domain as shown in Fig. 7.2.B. If the particle comes closer than $\mathscr{D}_{\text{min}}$ to the surface of the domain, then the domain is burst; a new position and orientation for the ES-complex (propagated as a single particle) is then drawn for the time elapsed since the construction of the domain. If the distance between the ES-complex and the other particle is now smaller than $2\mathscr{D}_{\text{min}}$, then the enzyme-substrate complex, which was propagated as a single particle, is replaced by an enzyme-substrate dimer, with the enzyme in the potential well of the appropriate patch of the substrate and with the same centre of mass and the orientation drawn from the FFS-BD pre-simulation; the three particles are then simulated with brute force BD. If the third particle is an enzyme particle that binds to the other free patch of the substrate, then the enzyme-substrate-enzyme (ESE) trimer is replaced by a single particle, and if possible subsequently put in a GFRD domain. On the other hand, if the third particle diffuses away, farther than $2\mathscr{D}_{\text{min}}$, then the enzyme-substrate dimer is again replaced by a single particle and a GFRD domain is built around it.

7.3.3. HANDLING DISSOCIATION, HOPPING AND PHOSPHORYLATION /DEPHOSPHORYLATION REACTION

Once an enzyme is bound to a substrate, one of three events can happen. The enzyme can either dissociate from the patch, phosphorylate/dephosphorylate the patch, or hop to the other patch, depending on the state of the patch (as mentioned, a kinase can only bind and hence hop to an unphosphorylated patch, while a phosphatase can only bind and hop to a phosphorylated patch). When the enzyme dissociates from the substrate, a new configuration is picked at random from the ensemble as generated in the BD-FFS pre-simulation. In this configuration as picked from the BD-FFS ensemble, the enzyme and substrate molecule not only have a certain relative position, but also a relative orientation with respective to each other; the position for the center-of-mass and the overall orientation of the particles in the complex are determined by the position and orientation of the complex (modeled as a single particle) just prior to the moment of dissociation. Here we capitalize on the idea that dissociation is a rare event: an event that happens rarely, yet rapidly when it does. Moreover, we assume it happens so fast, that we can take it to occur instantaneously, that is, no other event occur during the dissociation process. When the event is a hopping reaction, the enzyme particle ends up in the potential well corresponding to the other patch of the substrate; the relative position and orientation of the enzyme and substrate particle are again determined by the configuration picked from the BD-FFS pre-simulation and, as for dissociation, the overall position (center-of-mass) and orientation are determined by the position and orientation of the ES complex (which was modelled as a single particle) just prior to the event. Finally,

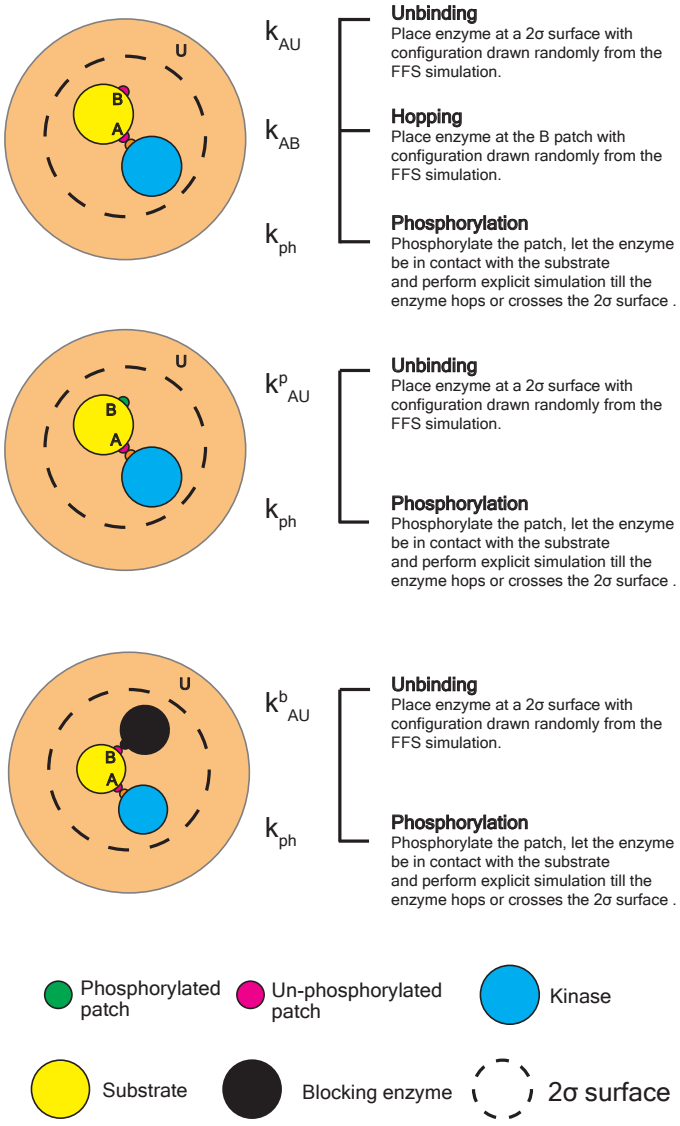


Figure 7.3: Handling dissociation/hopping/phosphorylation reactions in MD-GFRD. After an enzyme, say a kinase, binds to an unphosphorylated patch, the kinase can either dissociate, phosphorylate the given patch or depending on the state of the other patch, hop to the other patch. The figure shows the different rates that are calculated using BD-FFS pre-simulations. k_{AU}, $k^{\mathrm{p}}_{\mathrm{AU}}$ and $k^{\mathrm{b}}_{\mathrm{AU}}$ are the dissociation rates of the kinase starting in patch A and reaching the 2σ-interface before hopping to B or going back to A, when the other patch is unphosphorylated, phosphorylated and blocked respectively where σ is the radius of the particle. k_{AB} is the hopping rate of the kinase starting in patch A and reaching the patch B before diffusing to the 2σ-interface or going back to patch A, given that the patch B is unphosphorylated. k_{ph} is the rate of phosphorylation. After association, the next event times for all these events are drawn, the least of which will be the next event. If the next event is dissociation, then the kinase is placed at the 2σ-interface and if it is hopping the state of the patch of the substrate is changed to phosphorylated and the kinase is left in contact with the substrate. If the next event is hopping, the kinase is placed on the other patch. The rates for a phosphatase is taken to be the same as the above rates.

if the event is a phosphorylation or dephosphorylation reaction, then the ES complex (which was modeled as a single particle) is split into its products, the patch is phosphorylated/dephosphorylated and the enzyme is placed in contact with the same patch. As the patch interaction has now changed, the pair can dissociate and diffuse away spontaneously.

We define five types of rate constants: (i) Dissociation rate of an ES-complex, k_{AU} when the other patch is free and unphosphorylated if the dissociating enzyme is a kinase (or phosphorylated if the dissociating enzyme is a phosphatase). (ii) Dissociation rate of an ES-complex, $k^{\mathrm{p}}_{\mathrm{AU}}$ when the other patch is free and phosphorylated if the dissociating enzyme is a kinase (or unphosphorylated if the dissociating enzyme is a phosphatase). (iii) Dissociation rate of an ESE-complex, $k^{\mathrm{b}}_{\mathrm{AU}}$ when the other patch is blocked by another enzyme. (iv) Hopping rate for an ES-complex, k_{AB} to the other patch when its is free and unphosphorylated if the dissociating enzyme is a kinase or phosphorylated if the dissociating enzyme is a phosphatase. (v) Phosphorylation rate, k_{p}, of phosphorylating an unphosphorylated site, which equals the rate k_{dp} of dephosphorylating a phosphorylated site. The dissociation and the hopping rates are obtained from the BD-FFS pre-simulation. The rates and the events after association are illustrated in Fig. 7.3. These rate constants are used to determine the next event times for each of these events using Eq. 7.9. The event with the shortest tentative event time is the next event.

7.4. SYSTEM PARAMETERS

The system specific parameters of the simulation are as follows: The particle diameter is $\sigma = 5$nm, the time step $\delta t = 0.1$ns, the mass of the particle is $m = 50$kDa, the mass moment of inertia $M = \frac{8}{15} m\sigma^2$ the translational and rotational diffusion constants, are $D_t = 1\mu\mathrm{m}^2/\mathrm{s}$ and $D_r = 1.6 \times 10^5 \mathrm{rad}^2/\mathrm{s}$ for all particles, the translational and rotational friction coefficients are $\gamma = \frac{k_{\mathrm{B}}T}{D_t m}$ the $\Gamma = \frac{k_{\mathrm{B}}T}{D_r M}$ respectively, where $k_{\mathrm{B}} = 1.38 \times 10^{-23}\mathrm{JK}^{-1}$ is the Boltzmann constant and $T = 300$K is the temperature of the system. The dissociation and hopping rates obtained from the BD-FFD pre-simulation are $k_{\mathrm{AU}} = 0.73s^{-1}$, $k^{\mathrm{b}}_{\mathrm{AU}} = 1.1s^{-1}$, $k^{\mathrm{p}}_{\mathrm{AU}} = 1.4s^{-1}$ and $k_{\mathrm{AB}} = 0.7s^{-1}$. The phosphorylation and dephosphorylation rates are $k_{\mathrm{p}} = k_{\mathrm{dp}} = 15s^{-1}$.

We place into a cubic box with a side of 10σ, one substrate molecule and add a varying number of kinases N_{K} and phosphatases N_{P}. In this proof-of-principle simulation we vary N_{K} and N_{P} between 1 and 3.

7.5. RESULTS

We present the steady state input-output response of the MAPK signalling pathway. The response does not only depend on the rate constants, but also on the ratio of total enzyme to total substrate concentration: [K] + [P] vs [S_t] ratio. In Fig. 7.4 we present a proof-of-principle simulation for a system that is in the linear regime, where the enzyme concentration is much larger than that of the substrate: For both the kinase and the phosphatase we have 3 particles while we have only 1 substrate particle. We vary the concentration of the kinase over that of the phosphatase. The figure shows that the concentration of the active substrate, i.e. the doubly phosphorylated substrate, increases

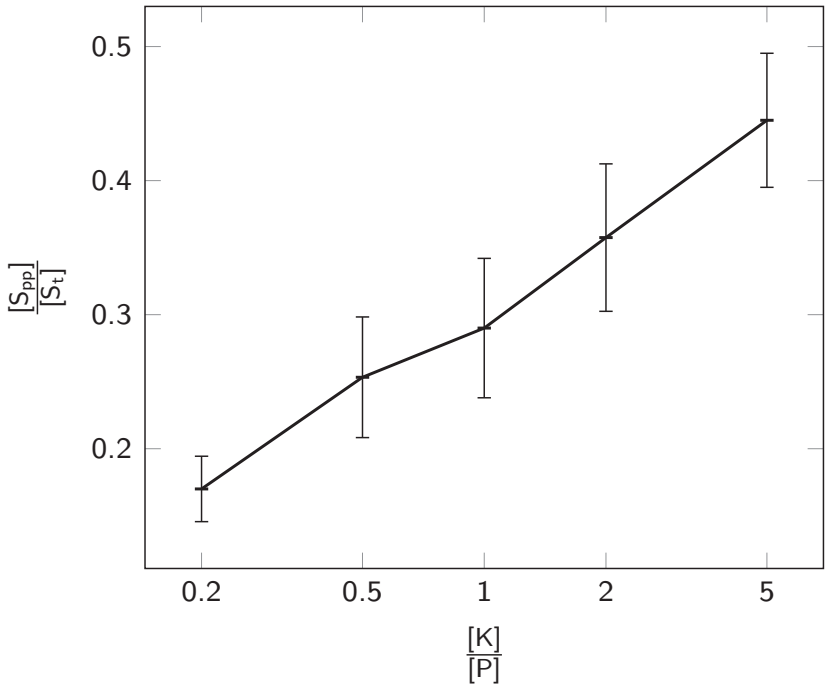


Figure 7.4: Steady state input-output relations for patches separated by 90°, with translational diffusion constant, $1\mu m^2/s$ and rotational diffusion constant 1.6×10^5. The simulation was performed with one substrate, three kinase and three phosphatase particles in a box of size 120σ, where $\sigma = 5nm$, is the radius of the particle. [K], [P], [S_{pp}] and [S_t]is the concentration of kinase, phosphatase, doubly phosphorylated substrate and the total number of substrate. In this case we see the enzymes phosphorylate the substrate processively, because after modifying one patch the enzyme rapidly rebinds to the other patch, instead of diffusing away. This is a proof-of-principle example to demonstrate that MD-GFRD can be used to simulate complex biochemical networks.

rather linearly with the ratio of the kinase to phosphatase concentration. This is is to be expected, given that the enzyme is in excess and the enzyme can rapidly rebind the substrate after modification of the first site, leading to pseudo-processivity.

7.6. OUTLOOK

In this paper we have presented one set of simulations to demonstrate that MD-GFRD can be used to simulate complex reaction-diffusion systems like biochemical networks. The aim is to simulate this system under more biologically realistic conditions, in particular more reasonable enzyme and substrate concentrations. Next, we would like to systematically study how the response depends on the orientational diffusion constant and the distance between the phosphorylation sites on the substrate. An interesting question will be whether these results can also be described (quasi) analytically, adopting the

theoretical approach of Gopich *et al.* and Ouldridge *et al.*. [12, 13]. Here the key quantity will be the rebinding probability or the hopping rate. More specifically, the question would be whether the response can be predicted from a modified chemical rate equation [12, 13], using a pre-computed hopping rate. Lastly, we point out that for this particular system it might be feasible to develop a custom-made code that, using precomputed rebinding probabilities and precomputed association propensities that depend on the phosphorylation states of the respective sites, can simulate this system very efficiently in a GFRD-like manner. However, MD-GFRD is exact and versatile, and can straightforwardly be applied to any reaction-diffusion system with patchy particles.

BIBLIOGRAPHY

[1] Takahashi K, Tănase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America* 107:2473–8.

[2] Paijmans J, Lubensky DK, ten Wolde PR (2017) A thermodynamically consistent model of the post-translational Kai circadian clock. *PLoS Computational Biology* 13:e1005415.

[3] Nash P, et al. (2001) Multisite phosphorylation of a cdk inhibitor sets a threshold for the onset of dna replication. *Nature* 414:514–521.

[4] Crabtree GR, Olson EN (2002) Nfat signaling: choreographing the social lives of cells. *Cell* 109 Suppl:S67–79.

[5] Miller P, Zhabotinsky AM, Lisman JE, Wang XJ (2005) The stability of a stochastic camkii switch: dependence on the number of enzyme molecules and protein turnover. *PLoS Biol* 3:e107.

[6] Chang L, Karin M (2001) Mammalian map kinase signalling cascades. *Nature* 410:37–40.

[7] Huang CYF, Ferrell, Jr. JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. 93:10078 – 10083.

[8] Ferrell, Jr. JE (1996) Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. 21:460–466.

[9] Ferrell JEJ, Bhatt RR (1997) Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *J Biol Chem* 272:19008–19016.

[10] Markevich NI, Hoek JB, Kholodenko BN (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol* 164:353–359.

[11] Aokia K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:12675–12680.

[12] Gopich IV, Szabo A (2013) Diffusion modifies the connectivity of kinetic schemes for multisite binding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 110:19784–19789.

[13] Ouldridge TE, ten Wolde PR (2014) The Robustness of Proofreading to Crowding-Induced Pseudo-Processivity in the MAPK Pathway. *Biophysj* 107:2425–2435.

[14] Vijaykumar A, Bolhuis PG, ten Wolde PR (2015) Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations. *The Journal of Chemical Physics* 143:214102.

[15] Vijaykumar A, Ouldridge TE, ten Wolde PR, Bolhuis PG (2017) Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics. *The Journal of Chemical Physics* 146:114106.

8

SUMMARY

In many reaction-diffusion processes, ranging from biochemical networks, catalysis, to complex selfassembly, the spatial distribution of the reactants and the stochastic character of their interactions are determine the macroscopic behavior of the system. While mean-field rate equations cannot describe such processes, the recently developed mesoscopic Green's Function Reaction Dynamics (GFRD) method enables efficient simulation at the particle level provided the microscopic dynamics can be integrated out. Yet, many processes exhibit non-trivial microscopic dynamics that can qualitatively change the macroscopic behavior, which requires an atomistic, microscopic description. In the first part of the thesis, i.e. chapters two and three, I describe our new simulation technique named Molecular Dynamics-GFRD (MD-GFRD), to simulate such reaction-diffusion system. In chapter two, I propose a novel approach that combines GFRD for simulating the system at the mesoscopic scale where particles are far apart, with a microscopic technique such as Langevin dynamics or Molecular Dynamics (MD), for simulating the system at the microscopic scale where reactants are in close proximity. This scheme determines the regions where the particles are close together, where the system is simulated using with a microscopic resolution using MD simulations. Regions where the particles are far apart allow using the more coarse grained simulation GFRD. Importantly, our algorithm switches adaptively between the two simulation techniques on the fly. The new multi-scale scheme MD-GFRD, is generic and can be used to efficiently simulate reaction-diffusion systems at the particle level. In chapter three, I extend MD-GFRD for simulating anisotropic particles with orientational diffusion. The next part of the thesis i.e. chapters four, five and six, deals with the derivation of microscopic expressions for the intrinsic and effective association and dissociation rate constants. Here, I address the role of the orientational dynamics, for particles interacting via anisotropic potentials. I systematically vary the strength and range of the interaction potential and investigate how these rates vary. I show how these rates can be computed using rare event simulations techniques. Using Forward Flux Sampling (FFS), I provide a generalized expression to calculate the transition rates, and use these expressions to calculate the rates for a system in which an enzyme binds to a substrate with two binding sites.

Chapter seven combines the two branches of the thesis to solve an important biological system, the MAPK signalling pathway. The association rates calculated in chapter six are plugged into MD-GFRD, such that the dissociation reaction need not be explicitly simulated. Our new simulation algorithm allows us to systematically study the response of the network as a function of the distance and angle between the binding sites, the rotational diffusion constant and the ratio of the substrate to the enzyme concentration.

9

SAMENVATTING

In veel reactie-diffusie processen, bijvoorbeeld biochemische netwerken, katalyse, en de zelfassemblage van moleculaire complexen, bepaalt de ruimtelijke ordering van de reagerende stoffen en het stochastisch karakter van hun onderlinge reacties in hoge mate het macroscopisch gedrag van het systeem. Omdat *mean-field* snelheidsvergelijkingen deze processen slecht beschrijven, is een simulatie op deeltjes niveau cruciaal. Simulaties waarbij elk deeltje expliciet beschreven wordt kosten echter veel tijd. Met het recent ontwikkelde *Green's Function Reaction Dynamics* (GFRD) is het wel mogelijk het systeem zeer efficiënt op deeltjes niveau te simuleren. Het werkt echter alleen als het microscopisch gedrag van de deeltjes grofkorrelig beschreven kan worden. Het macroscopisch gedrag van veel systemen hangt echter juist vaak af van het microscopisch gedrag van de onderliggende processen. Dit vereist een atomistische, microscopische beschrijving van de reacties. In hoofdstuk twee en drie beschrijf ik een nieuwe simulatie techniek met de naam Moleculaire Dynamica-GFRD (MD-GFRD), om reactie-diffusie systemen te simuleren. In hoofdstuk twee stel ik een nieuwe methode voor die twee bestaande simulatie technieken combineert: 1) GFRD, welke het systeem op mesoscopische schaal beschrijft als de deeltjes ver van elkaar verwijdert zijn, en 2) moleculaire dynamica, voor de microcopische beschrijving als de deeltjes dicht bij elkaar zijn. Het algortime bepaalt regionen waar de deeltjes dicht bij elkaar zijn en waar MD gebruikt moet worden, en de regionen waar slechts één deeltje voorkomt zodat GFRD voldoet. Het krachtige van ons algoritme is dat het naadloos de juiste simulatie metode kiest naargelang de lokale dichtheid van deeltjes. Ons nieuwe meerschalige algoritme MD-GFRD, is generiek en kan zeer efficiënt reactie-diffusie systemen op deeltjes niveau simuleren. In hoofdstuk drie breid ik MD-GFRD uit met de simulatie van anisotropische deeltjes en orientationele diffusie. In hoofdstuk vier, vijf en zes, leid ik de microscopische uitdrukkingen af voor de associatie en de bijbehorende dissociatie reactiesnelheidsconstanten. We berekenen zowel de intrinsieke en effectieve reactieconstanten. Hier behandelen we de rol van de oriëntatie in reacties tussen deeltjes met een anisotropische reactiepotentiaal. Ik varieer systematisch de sterkte en het berijk van de interactiepotentiaal en bestudeer hoe de reactiesnelheidsconstanten veranderen. Daarnaast laat ik zien hoe we deze con-

stanten kunnen bepalen met behulp van de *rare event* simulatie techniek Forward Flux Sampling (FFS). Ik leid een algemene uitdrukking af voor de reactiesnelheidsconstanten van een anisotropische potentiaal zodat we deze kunnen bepalen met behulp van FFS. In hoofdstuk zeven combineer ik alle technieken die ik ontwikkelt heb om een belangrijk biologisch systeem, de MAPK signaalweg, beter te begrijpen. De associatieconstanten die ik in hoofdstuk zes berekend heb gebruik ik in MD-GFRD zodat de dissociatie van een complex niet expliciet gesimuleerd hoeft te worden. Dit maakt de simulatie veel sneller. Ons nieuwe algortime maakt een systematische studie van het gedrag van een biochemisch netwerk mogelijk. Ik bestudeer de fosforylatie van een substraat in het MAPK systeem welke meerdere bindingsplekken voor bijvoorbeeld een kinsase heeft. Als functie van de afstand en de hoek tussen de bindingsplekken op het substraat, de orientationele diffusie constante en de ratio tussen de substraat-enzyme concentratie bepaal ik de hoe sterk het substraat gefosforyleerd is.

10

LIST OF PUBLICATIONS

Chapter 2
A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde - *Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations*, J. Chem. Phys. 143, 214102: 1-12 (2015).
`http://dx.doi.org/10.1063/1.4936254`
Chapter 3
A. Vijaykumar, T.E. Ouldridge, P.R. ten Wolde and P.G. Bolhuis, *Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics*, J. Chem. Phys. 146, 114106: 1-12 (2017).
`http://aip.scitation.org/doi/10.1063/1.4977515`
Chapter 4
A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde, *The intrinsic rate constants in diffusion-influenced reactions*, Farad. Discuss. 195, 421-441 (2016).
`http://dx.doi.org/10.1039/C6FD00104A`
Chapter 5
A. Vijaykumar, P.R. ten Wolde and P.G. Bolhuis, *The magnitude of the intrinsic rate constant: How deep can association reactions be in the diusion limited regime?*,
Under review in J. Chem. Phys.
Chapter 6
A. Vijaykumar, P.R. ten Wolde and P.G. Bolhuis, *Transition rates for proteins associating to substrates with multiple binding sites using a novel generic Forward Flux Sampling expression*,
Submitted to J. Chem. Phys.
Chapter 7
A. Vijaykumar, P.G. Bolhuis and P.R. ten Wolde, *Multi-scale simulations of multi-site phosphorylation in a MAPK signalling pathway*,
Manuscript in preparation


Cabo Verde
Rolling Stones
Radiker
The Doors
Ethiopia - Jazz
The Beatles
Goran Bregovic
Mulatu Astatke
Fatoş Cuaffman
Nina Simone


ENCOUNTER
ATTACK
40
Opposed by Attack:
Higher total (ships + card) wins.
Opposed by Negotiate:
Wins, but opponent collects compensation.
ENCOUNTER
40


30-5-2016
BANGALORE

11

ACKNOWLEDGEMENTS

Navigating the deepest chasms of the code in search of bugs, following the breadcrumbs left by std::cout, crossing vast areas of nothingness dotted with small oases of results, I have finally reached the end of my PhD. I have many people to thank for, since without them, this adventure would not have been half as exciting.

Pieter Rein, as a Star Wars enthusiast, growing up I always wanted a mentor like Yoda. Someone who not only taught me the tricks of the trade, but also shared experiential knowledge and was a constant source of inspiration. 'Dancing opens up your heart', is just one of many examples of your wise quotes. Your office was like a spa where I would come in stressed and walk out completely rejuvenated and with a confidence that I could tackle any scientific problem. The way you motivated me by telling your own stories from your PhD, where you made your favourite graph using an awesome script, is most appreciated. Also thank you for all the sailing trips and Sinterklaas dinners.

Peter, you are one of the most charismatic and charming professors out there. The way you present yourself and talk with people in a conference, reminded me of James Bond working himself through a casino. Your calm and cool attitude complimented with occasional but amazing sense of humour, made the meetings so much more enjoyable. Your exitement at the hint of some interesting results are testament enough to the passion you have toward science. I also am amazed by your ability to come up with extremely cool titles.

Joris, your selflessness puts me to shame. You are the kindest, most empathetic person I have ever known. You were my trusted travelling companion to Norway, Eastern Europe, Belgium, Eastern Turkey, Scotland, Andalucia and Cuba. You not only put up with all my last minute trip initiatives, but also indulged in them. Thanks for the godlike apple pies, the lessons in table manners, the poorest of jokes, bike rides, Lord of the rings quotes, for patiently listening to all my Indian stories (starring maga and his maga's friend), all the amazing sound effects, sharing your fascination of highway intersections, mathematica support, listening to all the nonsense I had to tell about paintings, bearing with the delays due to project-J when we travelled and thinking about the *key* question(s).

Martijn, when it comes to work ethics and organisation, we are on the opposite ends of the spectrum. Hopefully I have learnt a bit of both from you (ha-ha!!). Thanks for the tour of Nijmegen and the mountain biking trip. These four years you have been my sports partner for running, swimming, cycling, squash, surfing and bouldering and it will be hard if not impossible to find a replacement. I am still waiting for the day when you beat me at squash. I am astonished by your ability to very quickly make amazing food with limited ingredients. Thanks for all the times when you thoroughly brainstormed on my (non-) scientific problems and came up with amazing suggestions.

Chetan, thank you for organising all the biking trips along the North Sea in Norway and Sweden. Thank you for teaching me how to fold my underwear. Thanks for all the tips on driving, navigating and photography during our travels to Morocco, Portugal, Slovenia, Croatia and Montenegro.

One of the coolest places in Amsterdam is Levantkade 79. Michele your pastas were divine, and I still maintain that I have not had better pastas. Thanks for all the statistics discussions and the espressos made the Italian way. Jacopo thanks for sharing your passion for music, and playing all your latest compositions. Best part of course was the table tennis and the darts matches we played, with Oreste occasionally jumping over the table.

Tom, the two months I worked with you in London was by far the most productive I have been. Thanks for all the lessons in Statistical Mechanics, Mathematics and Moral Science. Your naval trivia, enthusiasm for cricket and rugby and love of history led to animated conversations. I am also glad that you are one of the committee members. You know that the Kohinoor belongs to us right?

Noreen, thanks for all the board game nights and bouldering tips and for putting up with rotten fruits crawling with animals that I used to leave in the office.

Stephen, thanks for making me understand that trying the local cuisine is one of the most important aspects of travelling, when we travelled to Turkey.

Joris, Martijn, Tom, Stephen and Noreen thanks for all the extremely nerdy board game nights, where it was often the case that we had to read rule books that were fatter than this thesis.

My knowledge of music was severely limited to Kannada and Hindi songs. But Ariana, you opened new worlds to me when you introduced me to western music. It all started with Parov Stellar's booty swing, next came AC-DC, Pink Floyd, Rolling Stones, Beatles, Goran Bregovic, Sigur Ross, and the list went on and on. Thanks for all the Mexican food that you made for lunch and all the extremely timely advice with respect to my thesis administration and for being an amazing office mate.

Arthur, thanks for all the discussions about TIS/FFS, the lunches at Polder, and the ground breaking, shockwave generating derivation that we did.

Nitish, thank you for indulging me in blunt Bangalore humour. Biking trips to Texel, Zaanse Schans, Muiden and of course the Amstel route was always enjoyable. I still think Masala Dose at Vidyarthi Bhavan is way better than the one in CTR, and that your Tamil sucks.

I was looking forward for every single board game night. Rakesh, you cut vegetables, Nitish, you 'cooked' and Ambuj, you were responsible for cleaning and I had the toughest job, which was putting up with all of you. We concluded every game night with a few

games of Cosmic Encounter, where Nitish often left extremely pissed off because Rakesh always had the most amazing cards (Cosmic/card zap, reinforcement or encounter 40).

A trip to Iceland is often very expensive, but Arjun and Shreyas, you took frugality to the next level. Sleeping in an information center, tent-ception, cooking at 3 a.m. in front of a huge glacier was all very memorable.

Yuliia, thanks for all the tips on watercolour painting and portrait proportions. Jurn, thanks for all the advice on consultancy companies and the inappropriate jokes.

I thank my group members Yao, Giulia, Harmen, Andrew, Chris, Sander, Verena, Ivan, David, Faidon and Alberto for all the helpful discussions.

Ivo, Ronald, Onno, Dylan, Martijn and Klas, thanks for making me part of the Table Tennis team. Ivo thanks for the coaching and all the racquets you gave me.

Kishore, Abhilash, Prapanch, Akhilesh, Varun and Pramod I am so glad I met you guys, it was fun growing up with you in Bangalore.

Appa, thanks for all the motivation and support and letting me choose a career that I like. Thanks for all the trips that you organised, in spite of the fact that I had no holidays left. Your skill as a photographer still amazes me. Your compositions can bring life to a still photograph, even now!! Amma, thanks for all the chapati nei sakkare, which got me through fifteen years of school. I thank you for putting our education above everything else and for the daily pep talks that motivated me to go to Germany. YB thanks for taking care of things back home, and making sure that I had peace of mind.

Even though you are thousands of kilometres away, Shreyanka, your presence in my life is very prominent. Thanks for the extremely graphic photos and videos of autopsies, which gave me nothing short of nightmares, and those of infants that you just delivered, which gave me a sense of pride. Thanks for being my art critique, model, interior designer, fashion adviser, grandmother, physician, dietitian, fitness trainer, financial planner, shrink and a patient audience sitting through all my practice presentations.