INAUGURAL – DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich - Mathematischen
Gesamtfakultät
der Ruprecht - Karls - Universität
Heidelberg

vorgelegt von
Dipl.-Phys. Jakob Schluttig
aus Leipzig

Tag der mündlichen Prüfung: 03.06.2009

Stochastic dynamics of protein assembly

Gutachter: Prof. Dr. Ulrich S. Schwarz

Prof. Dr. Heinz Horner

Zusammenfassung

Proteine sind an allen zellulären Prozessen beteiligt und agieren dabei in der Regel in Wechselwirkung mit anderen Proteinen. In dieser Arbeit wird ein Modell für die stochastische Dynamik von Proteinkomplexen untersucht, das sich für große Systeme und lange Zeiten eignet. Jedes Protein wird als Teilchen aufgefasst, welches an seiner Oberfläche reaktive Bereiche aufweist. Ein Überlapp solcher Bereiche von zwei Modellproteinen führt zur stochastischen Ausbildung einer Bindung, die dynamisch wieder aufbrechen kann. Die Anordnung und Kinetik der Bindungsstellen kann an konkrete biologische Systeme angepasst werden. Die Dynamik des Modellsystems wird mit einer Langevin-Gleichung im überdämpften Grenzfall beschrieben.

Zunächst werden drei verschiedene Detailstufen der Modellierung der bimolekularen Bindungsdynamik betrachtet. Die Zeit zum Erreichen einer reaktiven Relativposition zwischen zwei Proteinen und die Anzahl erfolgloser Annäherungen werden mit Computersimulationen und analytischen Rechnungen bestimmt. Der Effekt anisotroper Formen auf die Assemblierungsdynamik wird anhand von Ellipsoiden mit unterschiedlichen Aspektverhältnissen untersucht. Die relevante Zeitskale für anisotrope Diffusion wird analytisch bestimmt. Simulationen ergeben, dass die Annäherungsdauer im wesentlichen durch die unterschiedlichen Zugänglichkeit der reaktiven Bereiche bestimmt wird. Schließlich wird die Dynamik von Komplexen mit mehr als zwei Proteinen betrachtet. Es zeigt sich, dass die Transportprozesse zwischen Bindungsvorgängen nicht durch einfache stochastische Raten beschrieben werden können. Für die Virusassemblierung wird gezeigt, dass sie nur bei mittleren Dissoziationsraten erfolgreich verläuft.

Abstract

Proteins are involved in all cellular processes and typically act in concert with other proteins. Here we investigate a model for the stochastic dynamics of protein complexes which is especially suited to study large systems and long times. Each protein is modeled as a particle with reactive patches on its surface. An overlap of such patches of two model proteins leads to the stochastic formation of a bond, which also can dynamically dissociate. The positioning and kinetics of the patches can be adjusted to specific biological systems. The model dynamics is described by a Langevin equation in the overdamped limit.

We start by considering three different levels of detail for bimolecular encounter. The time to reach a reactive alignment of two model proteins and the number of unsuccessful approaches are determined by computer simulations and analytical calculations. The impact of anisotropic shapes on the assembly dynamics is investigated for ellipsoids with different aspect ratios. The relevant time scale for anisotropic diffusion is determined analytically. Simulations reveal that the time to encounter is mainly determined by the accessibility of the reactive patches. Finally the dynamics of complexes with more than two proteins is considered. We show that the transport processes between binding reactions cannot be described by effective stochastic rates. Our simulations also reveal that virus assembly is only effective at intermediate values of the dissociation rates.

Contents

Contents i

List of figures iii

List of symbols v

List of abbreviations viii

1 Introduction 1
- 1.1 Proteins – the main actors in the cell . . . 1
- 1.2 Protein–protein interaction . . . 4
- 1.3 Encounter complex . . . 5
- 1.4 Supramolecular complexes . . . 9
- 1.5 Overview . . . 12

2 Model and methods 17
- 2.1 Langevin equation approach . . . 17
- 2.2 First passage time problems . . . 23
  - 2.2.1 The Fokker-Planck equation . . . 23
  - 2.2.2 One-dimensional diffusion with periodic boundary conditions 26
  - 2.2.3 Diffusion between two absorbing spherical shells . . . 30
- 2.3 Evaluation of hydrodynamic properties . . . 33
  - 2.3.1 Hydrodynamic interaction in a system of many spheres . . . 34
  - 2.3.2 Simple example – the dumbbell . . . 39

3 Bimolecular encounter 45
- 3.1 Biologic examples at different levels of detail . . . 45
- 3.2 General scaling behavior . . . 48
- 3.3 Study of the encounter pathway . . . 52
  - 3.3.1 Three systems with different physico-chemical interface properties . . . 54
  - 3.3.2 Analytic approximations . . . 57
- 3.4 Discussion . . . 61

4 Influence of hydrodynamic anisotropy 65
4.1 From anisotropic to isotropic diffusion 65
4.1.1 Crossover in three dimensions 67
4.2 Brownian motion of rigid ellipsoidal particles 72
4.3 Influence of anisotropic shape on encounter rate 75
4.4 Discussion 79

5 Dynamics of clusters 81
5.1 Cluster of three homogeneously binding particles 81
5.2 Three-particle cluster with distinct geometry 85
5.2.1 Aligning bound model proteins in three dimensions 86
5.2.2 Network view of assembly – substates and transition dynamics 89
5.3 Modeling large complexes – Virus capsids 95
5.4 Discussion 98

Bibliography 101

Danksagung 115

List of Figures

1.1 Examples of protein complexes with different characteristics 2
1.2 Generic free energy landscape of protein-protein interaction 4
1.3 Assembly network of three-particle complex 10

2.1 Two particles on a line with periodic boundary conditions 26
2.2 First passage times of two-particle problem 28
2.3 First passage times averaged over initial conditions 29
2.4 Approximation of first passage time distribution 30
2.5 Two particles surrounded by spherical absorbing boundary 31
2.6 First passage times of spherical absorber problem 32
2.7 Analysis of the systematic error due to the Euler algorithm 33
2.8 Geometry of the dumbbell 39
2.9 Comparison of de la Torre method and exact results 43

3.1 Three bimolecular complexes at three levels of detail 46
3.2 Encounter first passage times and concentration dependence . . . 49
3.3 Frequency of first encounter for different particle numbers 50
3.4 Encounter rate dependence on patch size 51
3.5 Three alignment states and contact definition 52
3.6 Number of contacts distribution and resting and return times 53
3.7 Correlation plot of encounter rates and mean number of contacts . 56
3.8 Return to contact problem with approximate spherical symmetry . 58
3.9 Analytic expressions for contact resting and return times 61

4.1 Anisotropic random walk in two dimensions 66
4.2 Anisotropic to isotropic diffusion crossover in three dimensions . . 71
4.3 Diffusion coefficients and patch geometry of ellipsoid 73
4.4 Encounter rates of ellipsoid for different aspect ratios 76
4.5 Measure of patch accessibility 77
4.6 Relative deviation of encounter rates assuming isotropic diffusion . 78

5.1 Three particles on a line with periodic boundary conditions 82
5.2 Deviation of analytic and numeric mean encounter times 84
5.3 First passage times to fully bound system 85
5.4 Illustration of binding clusters in three dimensions 86

5.5 Relative rotation of binding clusters 88
5.6 Structure of triangular cluster and assembly network 89
5.7 Transition times and populations for different dissociation constants 92
5.8 Time series of cluster fractions 94
5.9 Local rules of T1 virus monomer and snapshots of self-assembly . 95
5.10 Time series of cluster fractions for T1 virus capsid 97

List of symbols

$\mathbf{A}(\mathbf{x}, t)$	Drift in Fokker-Planck equation, see Eq. (2.17) on page 24
$\mathbb{A}_i$	Matrix representation of cross product with vector $\mathbf{r}_i$, see Eq. (2.62) on page 35
$\mathbb{B}$	Cholesky decomposition of the mobility matrix on page 19
B_{ij}	Correction to electrostatic screening, see Eq. (2.14) on page 22
c	Protein concentration on page 47
D	Scalar translational diffusion coefficient, see Eq. (1.1) on page 5
$\mathbb{D}$	Diffusion matrix on page 18
$\mathcal{D}^t_{\parallel}, \ldots$	Approximate diffusion coefficients of the dumbbell, see Eq. (2.91) on page 41
Δ_{iso}	Relative deviation of encounter rates for anisotropic and isotropic diffusion, see Eq. (4.48) on page 79
$f(t)$	First passage time distribution, see Eq. (2.25) on page 25
$\overline{f}(t)$	First passage time distribution averaged over initial conditions, see Eq. (2.40) on page 28
$\overline{f}^{(1)}(t)$	First passage time distribution of first binding event for three-particle problem averaged over initial conditions, on page 82
$f^t_{\parallel}, \ldots$	Approximate normalized friction coefficients of the dumbbell, on page 43
f_V	Accessible patch volume fraction on page 77
$\mathbf{g}(\Delta t)$	Thermal noise contribution to the Euler update, see Eq. (2.5) on page 19
$G(t)$	Survival probability, see Eq. (2.24) on page 25

$\mathbf{j}(\mathbf{x}, t)$	Probability flux in continuity equation, see Eq. (2.16) on page 24
k	Encounter rate, i.e., encounter frequency scaled by concentration $k = \tilde{k}/c$ on page 48
$\tilde{k}$	Encounter frequency on page 48
k_a	Association rate from encounter to complex, see Eq. (1.4) on page 8
k_d	Dissociation rate from complex to encounter, see Eq. (1.4) on page 8
k_B	Boltzmann constant on page 18
κ	Inverse Debye screening length, see Eq. (2.14) on page 22
L	Edge length of the periodic boundary box, see Eq. (3.1) on page 47
$L_x, \ldots$	Diameters of ellipsoid on page 73
l_D	Debye screening length, see Eq. (2.14) on page 22
λ	Relative mobility of ellipsoid, see Eq. (4.40) on page 73
$\mathbb{M}$	Mobility matrix on page 18
$\mu^{tt}, \ldots$	Grand mobility matrices, see Eq. (2.58) on page 35
μ_{ij}^{tt}	Translational part of the mobility matrices due to pair-interaction, see Eq. (2.68) on page 36
N	Number of contacts before encounter on page 52
N_A	Avogadro constant on page 47
$p(\mathbf{x}, t)$	Probability density on page 23
$\hat{\mathbf{p}}$	Dipole of a protein on page 22
ψ_z	Angle defining location of patch on ellipsoid on page 73
$\hat{q}$	Total charge of a protein on page 22
q_i	Effective model charge i on page 22
R	Radius of model protein on page 22
r	Radius of reactive patch on page 51
r_p	Distance of effective dipole charges from center of mass, on page 22
$\mathbf{r}_{DO}$	Location of the center of diffusion, see Eq. (2.73) on page 37

T	Mean first passage time, see Eq. (2.26) on page 25
$\overline{T}$	Mean first passage time averaged over initial conditions, see Eq. (2.39) on page 28
$T^{(1)}$	Mean first passage time to first binding event for three-particle problem on page 82
T_{off}	Return time to contact on page 52
T_{on}	Resting time in contact on page 52
T_a	Ambient temperature on page 18
Δt	Finite time step, see Eq. (2.5) on page 19
Δt_{ad}	Variable time step, see Eq. (2.9) on page 20
Δt_{min}	Lower boundary for the time step on page 20
$\langle \mathbf{x}(t)^2 \rangle$	Mean square displacement for anisotropic diffusion, see Eq. (4.30) on page 71
ξ	Aspect ratio of ellipsoid on page 73
$\mathbf{\Xi}_O^{tr}, \ldots$	Friction matrices of the rigid aggregate with respect to point of reference O, see Eq. (2.66) on page 36 and Eq. (2.72) on page 37
$\mathbf{\Xi}_{\mathrm{corr}}^{rr}$	Correction to rotational friction matrix, see Eq. (2.78) on page 38
ζ	Friction matrix, see Eq. (2.1) on page 17
$\zeta^{tt}, \ldots$	Grand friction matrices, see Eq. (2.58) on page 35
$\zeta_{ij}^{tt}, \ldots$	Friction matrices due to pair-interaction, see Eq. (2.56) on page 35 and Eq. (2.57) on page 35

List of abbreviations

$\mathcal{A}1 - \mathcal{A}3$	Alignment states	page 52
$\mathcal{C}1 - \mathcal{C}6$	Assembly substates	page 89
DSM	Dipolar sphere model	page 21
EM	Electron microscopy	page 1
FPE	Fokker-Planck equation	page 23
FPT	First passage time	page 23
FRET	Fluorescence resonance energy transfer	page 3
$\mathcal{G}1 - \mathcal{G}3$	Locations of patches on ellipsoids	page 75
GFRD	Green's function reaction dynamics	page 20
LJ	Lennard-Jones (potential)	page 11
$\mathcal{M}1 - \mathcal{M}3$	Levels of detail in our model	page 46
MFPT	Mean first passage time	page 23
NMR	Nuclear magnetic resonance	page 1
PDB	Protein data bank	page 34
$\mathcal{S}1 - \mathcal{S}3$	Three bimolecular systems	page 46

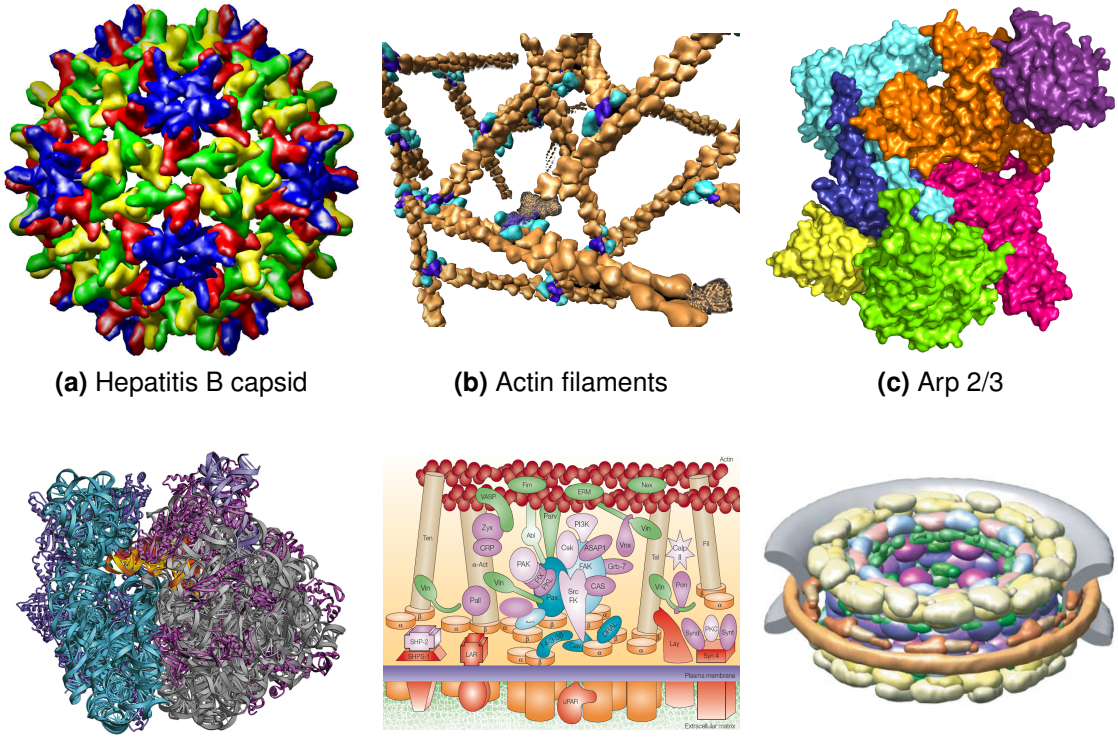
Chapter 1

Introduction

1.1 Proteins – the main actors in the cell

Proteins are a large class of highly specialized macromolecules performing essential functions in biological systems, such as processing of genetic information, signal transduction, buildup of cellular structure like the cytoskeleton and transport of material [6, 154]. Structurally, proteins are polymers build up as sequences of 50–3000 amino acid residues linked by peptide bonds. All proteins are composed of 20 different types of amino acids. The particular sequence of amino acids in a protein, also referred to as *primary structure*, is encoded in the genes. This sequence is responsible for the formation of a stable and unique native conformation. A whole cell machinery, partially consisting of proteins itself, transcribes DNA into mRNA and synthesizes proteins according to the mRNA. The three-dimensional structure of the protein itself, also called the *tertiary structure*, determines the biological function of a protein. Since 2001, the whole human genome is sequenced [151]. In principle, this should imply a deep understanding of the majority of cellular processes. In practice, the conceptual step from the genetic sequence to the multitude of molecular species and their concerted function inside the cell is huge and poorly understood. Regarding single proteins a lot of progress has been made concerning the problem of protein structure determination from the primary structure. Over the past decades a large number of tertiary structures have been discovered by increasingly powerful methods such as x-ray crystallography [80], nuclear magnetic resonance (NMR) spectroscopy [161] or electron microscopy (EM) [56]. The growing amount of this information leads to an improvement of bioinformatics approaches to structure prediction like homology modeling and protein threading [168]. Altogether, this development marks a big step towards the understanding of the biological function of proteins.

However, in recent years it has been noticed that proteins are mainly functional in complexes [19, 70, 81]. In fact, about half of all cellular proteins are assumed to be part of macromolecular complexes at various times [42, 59, 60]. The average oligomeric state of soluble proteins in *Escherichia coli* is four, with 15% forming high-order assemblies. In the biochemistry community, biologically functional complexes are usually termed *assemblies*. In contrast, the term *aggregation* refers to pathologic clustering and plaque formation as found in several



(a) Hepatitis B capsid **(b)** Actin filaments **(c)** Arp 2/3

(d) 70S Ribosome **(e)** Focal adhesion **(f)** Nuclear pore complex

Figure 1.1: (a) Hepatitis B virus capsid [162]. The whole capsid is built from only one type of monomers, which are shown in different gray scales according to their alignment in the icosahedral face. (b) Artificial tree of actin filaments, taken from UCSF Chimera Gallery (2004). (c) Cartoon representation of the 70S Ribosome [164]. (d) Arp 2/3 complex [119]. It is built up from seven different proteins each shown in a unique gray scale in the image. (e) Sketch of a focal adhesion [61]. (f) Nuclear pore complex [5].

neurodegenerative diseases such as Alzheimer or Parkinson. The process of formation of assemblies out of the individual proteins usually occurs without external control and is therefore called *self-assembly*. Functional protein complexes can have various characteristics and take over all kinds of different functions. Here, we want to define three different types of complexes without claiming completeness of categorization. First, there are supramolecular complexes functioning as *architectural scaffolds*. Prominent examples are virus capsids (see Fig. 1.1a), actin filaments (see Fig. 1.1b) or clathrin coats. These are often made up of a large number of copies of only a small number of different components. Therefore, they are typically highly regular and symmetric [14, 102, 165]. Apart from a particular growth dynamics like actin polymerization and depolymerization, such complexes are typically very stable, i.e., their structure is well determined and rather permanent. Different characteristics are found for *biomolecular machines* like the ribosome (see Fig. 1.1d), the spliceosome, photosynthetic reaction centers or the Arp2/3 complex (see Fig. 1.1c). The respective compounds are more heterogeneous and typically consist of an average number of $\sim$10 different protein species.

Some of the biomolecular machines have a defined dynamic behavior to perform their tasks, such as protein synthesis in ribosomes or the photophosphorylation in bacterial reaction centers. Finally, there are large *open complexes* like adhesion clusters (see Fig. 1.1e), transcription complexes, motor complexes or the nuclear pore complex (see Fig. 1.1f). These can consist of up to millions of proteins out of hundreds of different species as in case of the focal adhesion complex. Focal adhesions are the main sites of cell attachment to the extracellular matrix and control many cellular processes such as proliferation and migration. They have a particularly rich dynamics because the adhesion is continuously adjusting to changes in the environment and thus there is large turn-over in the system. In contrast, once formed the nuclear pore complex is rather stable, but during the cell division the huge complex is completely disassembled and built up again in both daughter cells. These considerations show that unlike for most single proteins dynamics is very important for the function of protein complexes [42].

Regarding the composition and structure determination of protein complexes a number of experimental and computational techniques have been adopted from single protein investigations or newly developed in the field of *proteomics* [106, 122]. These methods range from x-ray crystallography, NMR spectroscopy, and electron tomography (which are able to give direct information about assembly shape and structure) through fluorescence resonance energy transfer (FRET) microscopy, mass spectroscopy, and gene/protein arrays (providing subunit proximity measures) to theoretical bioinformatics and docking approaches (predicting subunit-subunit contacts). An impressive example of the power of these techniques was the recent solution of the structure of the whole nuclear pore complex containing 456 proteins by the integration of a broad range of the available methods, as well as reasoning due of geometric constraints and further optimization [4, 5]. In contrast to the rapid advances in structure determination the analysis of realtime protein complex dynamics in experiment is still unsolved. Although FRET microscopy is able to observe protein encounters *in vivo* and in realtime, not *all* proteins can be labeled, so that a lot of encounters are possibly missed and the actual time lapse is misinterpreted. However, mass spectrometry methods have been successfully used to unravel dynamics of complexes [132]. For example, in a recent study with nanoelectrospray mass spectrometry the subunit exchange in dodecamers of small heat shock proteins has been explored [105]. Still, experimental data in this field is rare, giving strong biological motivation for various modeling approaches.

From a physics point of view, protein assembly is a many-particle problem. Such systems are at the core of statistical mechanics. Traditionally, research in physics has focused on systems of particles with non-specific interactions. Colloidal sciences mainly investigated assembly into bulk thermodynamic phases like liquid crystals, gels and attractive or repulsive glasses [130]. Recently interest has shifted towards colloids or nanoparticles with more specific assembly properties. One of the motivations for this development is the attempt to use the increasing knowledge about protein complexes to design artificial systems with similar functionalities [64]. Experimental techniques have become available to create particles of various shapes and with distinct functional sites at the surfaces [94].

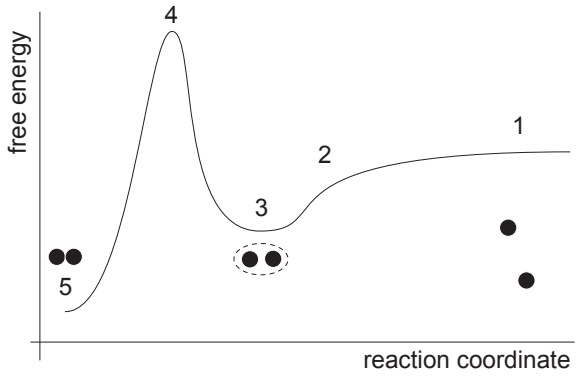


Figure 1.2: Schematic effective free energy landscape of protein association. 1) free diffusion, 2) electrostatic steering region, 3) encounter complex, 4) barrier due to desolvation and other effects, 5) final complex.

Gracias et al. succeeded in building mm-sized particles which self-assembled into helical structures [72]. Blum et al. placed gold nanoparticles of different size at distinct positions of a Cowpea Mosaic Virus [27]. This procedure shows that in principle it is possible to functionalize virus capsids and thus create very small building blocks with well-determined features. Also on the theory side the physics of solutions of colloids with anisotropic interactions is investigated. Zhang et al. performed computer simulations where particles with discrete interaction sites self-assembled into chains, sheets, rings and other structures [169]. The phase transition to sheets is found to be of second order, while chains form through a first order transition. Analytic solutions for the fluid-fluid transition of sticky, patchy hard spheres have been found by Fantoni et al. [55]. The results agree with corresponding Monte Carlo simulations. In summary, the research effort invested into comparable questions is large and the cited references are only examples for the current interest in the field.

1.2 Protein–protein interaction

In contrast to the dynamics of larger protein complexes, which is still relatively unexplored, a lot of both experimental and theoretical research effort has been invested into the understanding of bimolecular protein interactions. These studies revealed that on the one hand the details of protein interaction are diverse and complex, while on the other hand they show particular similarities, especially on scales larger than atomistic. Fig. 1.2 shows a generic free energy landscape of the interaction of two proteins. The reduction to a single reaction coordinate is a strong simplification but reflects the fact that protein association is a highly controlled process, characterized by one most probable reaction trajectory. Initially the binding partners undergo pure diffusion (1). Although Coulombic interactions are typically considered as *long-ranged*, in biological contexts they are screened on the scale of the Debye length of about 1nm because of the high concentration of ions in the cytoplasm under physiological conditions. After reaching a certain

proximity, the proteins are steered towards each other (2), usually by electrostatic and possibly hydrodynamic forces. The latter are assumed to have a rather weak impact [15] and are thus neglected in this work. The electrostatic steering tilts the free energy landscape towards the final complex. For even smaller distances, a number of complicated effects at atomistic scale result in a final free energy barrier (4) [2]. For example, proteins typically present hydrophilic aminoacids to the surrounding solvent and are thus covered by a layer of water molecules which has to be removed before they can get into close contact. This phenomenon is called *dehydration*. Also, in some cases the binding interfaces fit rather tight. Thus, certain structural fluctuations are necessary to permit the attachment and one has to account for conformational dynamics [90]. Furthermore, the configuration space is typically getting much more narrow during the binding process and thus the entropy is strongly decreased, which in turn increases the free energy. The interplay between the barrier (4) and the steering effects (2) leads to the formation of a local free energy minimum (3) – the *encounter complex*. Given the simplifying considerations leading to this concept it is clear that an encounter complex does not exist in every case. However, it has been confirmed for several important cases. Moreover, it is a very helpful concept to model protein interaction networks. The encounter complex can be considered as a kinetic concept in the first place rather than having a particular structural representation. As the encounter complex is typically transient, its detection is complicated, especially in experiment. However, recent studies proved its existence in a particular system and even unraveled particular structural aspects [79]. The height of the *dissociation barrier* (5) $\rightarrow$ (3) is believed to be mainly controlled by short-ranged interactions like hydrogen bonding and van der Waals forces. In this work we use the encounter complex as the conceptual starting point to define a stochastic dynamic model for protein assembly. For a recent review see Ref. [127].

1.3 Encounter complex

Analytical and computational considerations with respect to the encounter complex have a long tradition [50, 57, 127]. Early works attempted to find mathematical descriptions for the encounter step, i.e., for the transport part of the reaction. One prominent result is the rate of encounter between a colloid of finite size and an ensemble of small colloids derived by Smoluchowski as a solution of the corresponding diffusion equation [139]:

$$k_+ = 4\pi D r_0 \ , \tag{1.1}$$

where D denotes the sum of the isotropic diffusion coefficients of the two types of colloids $D = D_1 + D_2$ and r_0 is the capture radius or in other words the radius of the collision cross section of the two colloids $r_0 = r_1 + r_2$. However, because protein-protein binding is anisotropic, this formula can be used only as a reference framework for our purpose. Debye calculated reaction rates in ionic solutions [44]. Eigen did the first step towards the consideration of the encounter complex in a biological context, particularly enzyme physics [47]. He discussed the classic work

by Debye [43] and the two limits of pure electrostatics [91, 103, 104] and pure geometry [139]. Berg and Purcell introduced the standard model for this field (for ligand–cell binding rather than ligand–receptor binding) [22]: ligands diffuse to a sphere coated with disk-like receptor patches and are immediately captured upon encounter. Interestingly, for typical values from cell receptor applications, already a very low surface coverage ($\sim 10^{-3}$) leads to nearly optimal outcome. Later, Zwanzig discussed cooperative effects between the receptor patches [176] and derived a correction to the Berg and Purcell result, which matched simulations by Northrup [98]. Bell considered a two step binding process and the possibility of dissociation [21]. Consequently, DeLisi and Wiegel discussed the Berg-Purcell model with a finite reaction rate and accounting for electrostatic interactions [45]. They claim that, although the particular kinetics of association and dissociation can be affected, the equilibrium properties remain the same. Shoup and Szabo used the concept of a radiation boundary condition to model the formation of the final complex from the encounter state in a mathematically rigorous way [136]. Their treatment includes electrostatic interactions and is not restricted to the diffusion-limited case. For the latter, however, they are able to reproduce the results of Berg and Purcell. In the following years, Goldstein and Thompson worked out more details [68, 69, 89, 160].

In many cases, experimental rates were found to be larger than predicted by the theoretical work. As a consequence, more specific properties like the particular geometry of the receptor patches were considered in the models. Shoup and Szabo discussed the impact of orientation constraints and rotational diffusion. They find that the latter can strongly decrease the encounter time of a ligand and a receptor patch like in the Berg-Purcell model [135]. Northrup started to use computer simulations to study protein-protein association [99]. He showed that the rate enhancement seems to be caused by an entrapment of the encounter complex, which allows for sampling a large number of alignments without leaving the encounter state [100]. Barzykin and Shushin claimed that disk-like patches as used in the Berg-Purcell model lead to substantially lower reaction rates than the use of hemispherical patches [20]. In a further publication the same authors suggest that anisotropic shapes of molecules can enhance the reaction rate [137]. The importance of electrostatic interactions for long-ranged attraction was emphasized by Brownian Dynamics simulations of protein-protein encounter [49, 99, 155, 171]. If atomic structure is taken into account, then successful encounters are defined by simultaneous fulfillment of two to three distance conditions between opposing residues on the two surfaces [58]. Recently Korn and Schwarz [85, 86, 87] used a purely geometrical interpretation of the encounter complex to study the efficiency of cell adhesion in hydrodynamic flow, where convection competes with diffusion. Erdmann and Schwarz [52] used the concept of a position-dependent rebinding rate to study the role of cell-substrate distance in cell adhesion. Brownian Dynamics have also been used for the simulation of dense systems, e.g. by Bicout and Field who studied a cellular “soup” containing ribosomes, proteins and tRNA molecules [26], or recently by Elcock and coworkers who studied a crowded cytosol for 10μs length [93].

Schlosshauer and Baker extended the work of Shoup and derived the bind-

ing rate for two spherical molecules which bind at hemispherical reactive patches with a finite reaction rate [123]. In a recent study, Alsallaq and Zhou [11] again extended this 'hemisphere' model by introducing a 'crater' model consisting of a spherical protein with a crater to which another spherical protein fits snugly. There, the formation of a stereospecific complex was disfavored by the loss of translational and rotational freedom. Small translations and rotations between the protein subunits destroyed the interactions, leading to a sharp transition between the bound and the unbound state. The energy landscape was described as funnel-like, with the deep well of the bound state surrounded by a broad shallow basin.

Several recent studies have addressed the encounter complex in great detail. Miyashita et al. [95] investigated the effect of electrostatic interactions on the binding reaction between cytochrome c2 and a bacterial reaction center. The mechanism involved an encounter complex stabilized by electrostatic interactions, followed by a transition state similar to those found by Zhou [11], leading to the bound complex active in electron transfer. The study involved determination of a set of transition state structures by fitting experimental kinetic data over a wide range of protein-protein configurations. The transition state ensemble, obtained from structures having the highest correlation coefficients in comparison with the experimental data, had the cytochrome displaced by about 10Å from its position in the x-ray crystal structure. The observed similarity between the structures of the encounter state, transition state, and bound complex accounted for the rapid rate of association responsible for fast diffusion-controlled electron transfer. Spaar and Helms [141] used Brownian Dynamics simulations in order to study the association of barnase and barstar. The individual positions and orientations of one protein relative to the other were interpreted as a probability distribution allowing the calculation of the entropy landscape. The free energy landscape was obtained by summing the electrostatic, desolvation, and entropy contributions. A characteristic minimum at about 10Å distance between the two binding patches denoted the position of the encounter complex. Recently, it has become possible to investigate the nature of transient intermediates under equilibrium conditions via paramagnetic relaxation enhancement [79]. Consequently, this technique was applied to three different complexes from the bacterial phosphotransferase system and the results were compared to replica exchange simulations [82]. Indeed, both experiment and simulation consistently showed a relative population of about $\sim 10\%$ nonspecific complexes representing a combined landscape of transients close to the specific complex.

This picture of protein association suggests to view it as a two step process. This concept extends the original model of receptor (R) ligand (L) interaction, where the final complex (C) associates or dissociates by single kinetic rates k_f (forward) and k_r (reverse) respectively [92]:

$$\mathrm{R} + \mathrm{L} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{C} \; . \tag{1.2}$$

The corresponding kinetic equation describes the time evolution of the concentration of the complex C with respect to the concentrations of receptors R and ligands L:

$$\dot{C} = k_f RL - k_r C \; . \tag{1.3}$$

This simple form is only valid if there is a large reservoir of receptors and ligands and thus R and L are not decreased by complex formation. A simple dimension analysis in Eq. (1.3) reveals that k_r is given as a frequency $[k_r] = 1/\mathrm{s} = \mathrm{Hz}$ while k_f must have units per time per concentration $[k_f] = 1/\mathrm{sM}$, where M means molar $\mathrm{M} = \mathrm{mol/l}$. Considering the steady state $\dot{C} \equiv 0$, the *equilibrium dissociation constant* $K_D = k_r/k_f$ can be used to express the complex concentration $C = RL/K_D$. Bell first discussed the case where the reaction forming the final complex is preceeded by an encounter step E in the context of cell-cell adhesion [21]:

$$\mathrm{R} + \mathrm{L} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{E} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{C} \; . \tag{1.4}$$

The kinetic description now has to deal at least with the concentrations of encounter complexes E and finally bound complexes C:

$$\dot{E} = k_+ RL + k_d C - (k_- + k_a) E \; , \tag{1.5}$$

$$\dot{C} = k_a E - k_d C = k_f RL - k_r C \; . \tag{1.6}$$

As the encounter complex is thought to be transient, E will be typically small and Eq. (1.5) can be considered to be in a quasi-steady state $\dot{E} = 0$. In this case it is possible to calculated effective rates k_f and k_r related to an overall reaction of the type of Eq. (1.2):

$$k_f = \frac{k_+ k_a}{k_- + k_a} \; , \qquad k_r = \frac{k_- k_d}{k_- + k_a} \; . \tag{1.7}$$

For sequential stochastic processes it is common to identify the *rate limiting* step, i.e., the part of the sequence having the biggest impact on the overall kinetics. For the case presented here two limits are important as discussed by Bell in his original work. If the reaction step happens much quicker than the dissociation of the encounter $k_- \ll k_a$, we have:

$$k_f \approx k_+ \; , \qquad k_r \approx k_- \frac{k_d}{k_a} \; , \tag{1.8}$$

which means that the overall process is *diffusion controlled*. Conversely, $k_- \gg k_a$ describes the *reaction controlled* limit:

$$k_f \approx k_a \frac{k_+}{k_-} \; , \qquad k_r \approx k_d \; . \tag{1.9}$$

Indeed, many protein-protein interactions are found to support the diffusion controlled limit. A phenomenological biological explanation is that for low protein concentrations the encounter rate due to diffusion will be typically small $k_+ \ll k_-$. In the reaction limited case this would basically mean that k_f is the product of two small quantities divided by one large one – which gives an extremely small result. Therefore, it seems reasonable that protein reactions might be mainly diffusion controlled to stay functional at low concentrations.

It is important to note that the considerations in this section presume each step in the reaction scheme to be a Poisson process, which is defined by a single stochastic rate. However, this is *not* a priori clear and must actually be seen as a convenient simplification. In particular, the reaction step, i.e., from E to C does only behave Poisson-like if the final barrier (state 4 in Fig. 1.2) is steep and narrow, which again is an assumption that has to be proven by detailed investigations [3].

However, despite the extensive work on bimolecular encounter, the stochastic dynamics of larger complexes is still a largely unexplored subject.

1.4 Supramolecular complexes

The separation of the association of two proteins into an encounter and a reaction step can be utilized to build up stochastic networks of encounter and reaction processes to reflect the formation of supramolecular complexes with more than two building units as depicted in Fig. 1.3. Assuming all processes to be of Poisson type, these networks can then be treated with standard methods from stochastic dynamics [78]. The second advantage of the concept of an encounter complex is that encounter and reaction processes happen on very different scales, and thus the encounter complex provides a suitable crossover scale for multiscale modeling. The diffusional encounter can be described in a more coarse-grained manner compared to the reaction step, where an atomistic treatment of many short-ranged interactions is necessary.

In Fig. 1.3 we schematically illustrate how the concepts of an encounter complex can be used to address the dynamics of complexes. It has become apparent that for certain systems spatial effects like local precursor dilution after reactions can be important. This is especially the case for low concentrations, where there is rather a finite number of copies of a particular protein in the cell than a homogeneous concentration. In general the level of detail required to implement the regarding effects in a model does not allow for analytic treatment. Consequently, progress has been made particularly in terms of computer simulations on reaction-diffusion problems. As the number of more or less specific software approaches addressing this issue is rapidly growing we want to exemplarily name only a few. `E-CELL` is a sophisticated project that integrates a variety of algorithms and common coarse-graining techniques as well as many of the available information about the genome and protein interactions in yeast in a simulation of the whole yeast cell [146]. A software which is mainly concerned with a detailed representation of spatial constraints to diffusion in the extracellular environment of, e.g., neuronal synapses is `MCell` (see Ref. [37] and references therein). The

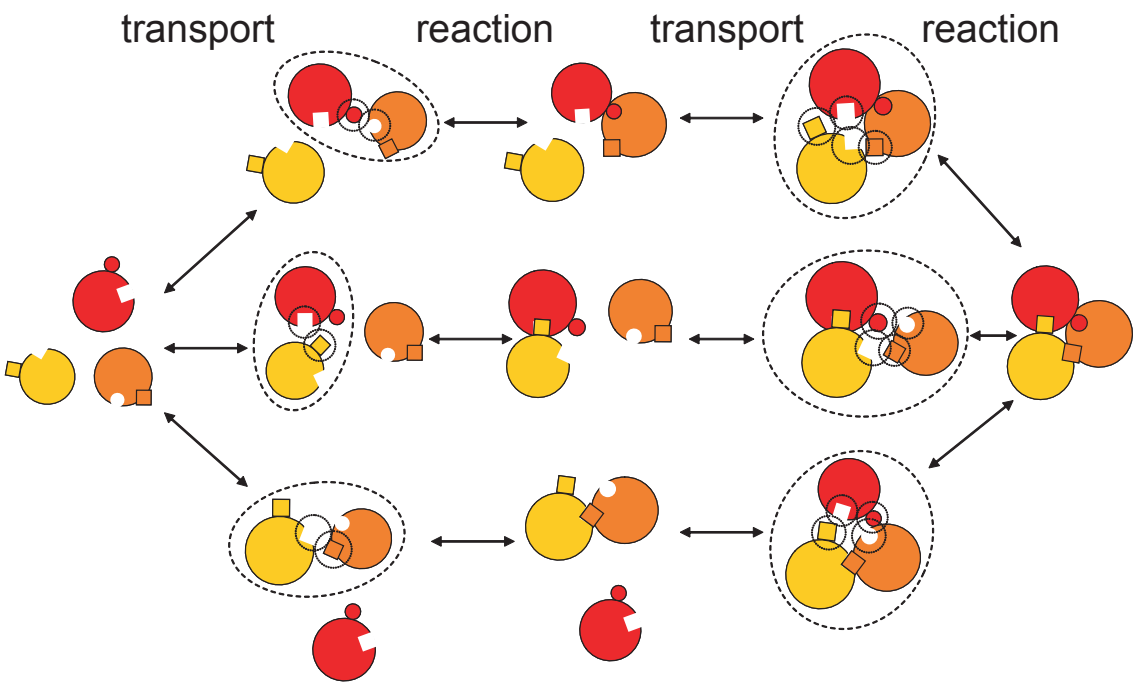


Figure 1.3: Generic network representation of the association of a complex of three different proteins. Each protein-protein bond is the product of a transport step leading to the formation of an encounter complex (denoted with dashed lines) and a reaction step. Considering unique proteins there are a total of 11 possible states of the complex including the completely unbound state as well as possible encounter or partially formed complexes. The possible pathways due to reversible transport or reaction processes are indicated by arrows.

lattice based simulation `MesoRD` effectively models diffusion from one lattice cell to the next as a special reaction event occurring at some stochastic rate adjusted to lattice resolution and diffusion constant. With this approach, the authors discovered noise as the main reason for differing Min oscillation patterns in several *Escherichia coli* mutants [54]. While this work utilizes an extended version of the Gillespie algorithm [62, 63] for quick simulations referred to as the *next subvolume method*, the particular position of a single particle is only known in terms of the lattice cell it is contained in. A promising development is therefore the *Green's function reaction dynamics* [149, 150]. In this off-lattice algorithm, the time of the next reaction of a number of freely and isotropically moving particles with a certain reactivity is calculated. This makes it possible to overleap phases of free diffusion in the numeric scheme which makes it very effective. By a simple model of gene expression the authors show that Green's function reaction dynamics can be up to five orders of magnitude faster than conventional techniques.

While the aforementioned works consider the local distribution of monomers, the study of specific three-dimensional assemblies requires even more detailed approaches. Here, the information about the number and kind of monomers in a particular cluster has to be extended by their specific alignment in the complex. Also, diffusion will not remain isotropic because especially growing clusters can

have shapes strongly differing from spherical. A class of prominent examples for large macromolecular structures are virus capsids (see Fig. 1.1a). The dynamics of their self-assembly has been the subject of several studies. Capsids are typically built from only one or a small number of distinct protein types, sometimes with different possible conformations. Therefore, they are in most cases highly regular and symmetric [14, 102, 165]. Capsid proteins have been reported to be able to passively self-assemble into capsids with a defined number of proteins in *in vitro* experiments [175]. The self-assembly can be viewed in the framework of a nucleation process, i.e., there is a lag phase until a critical concentration of stable nuclei has formed, and the growth of the full capsids is fast afterwards. However, the nucleation of virus capsids is qualitatively different from classic nucleation systems like linear polymers (actin) or crystals. As the number of monomers in the final complex is fixed to some definite number in the range of tens or hundreds, a large number of capsids will form which in turn also requires a large number of nuclei [173]. Thus, different approaches in terms of nucleation theory have been developed to describe the specific problem of capsid assembly [166]. To gather information about structural aspects of the self-assembly process and find the particular structure of partially formed capsids which are important for the assembly pathway, more detailed studies are required. Berger et al. introduced a theory of local rules [23] which revealed that different types of virus like structures can be encoded in tiny sets of rules that define bond formation at each monomer locally. The authors showed in a more or less static growth study that perturbations of the rules or mistakenly bound monomers may strongly decreased the probability of the formation of a closed capsid or even prohibit it. This prediction agrees with the experimental finding that a small molecule can inhibit and misdirect the assembly of Hepatitis B virus capsids [174]. Later, the theory was extended by a set of dynamic rules in terms of harmonic potentials and several sample trajectories of the extended model were generated [129]. Another example for the generality of the local rules approach is its successful application to polyomavirus polymorphic capsid assemblies [128]. Wilber et al. distinguish phases of assembly of many closed shells and aggregation of large clusters of particles with reactive regions built up analogously to the theory of local rules by the use of effective Lennard-Jones (LJ) interactions in a Monte Carlo study [159]. Rapaport used a different approach in a constant temperature Molecular Dynamics simulation with specifically designed subunits built up from spheres and with effective LJ-like interactions [116]. In a recent continuation of his work [117], more detailed simulations including an explicit solvent representation revealed that the highest possible number of bonds in partially formed clusters is preferred in the model due to the implied energetics. Furthermore, it is shown that reversibility of bond formation is of particular importance to prevent kinetic trapping of the system with many unfinished capsids which cannot be connected to the desired final structure. This finding agrees well with results of Hagan and Chandler [74], who studied a system of spherical particles with another type of effective LJ potential favoring cluster formation similar to the local rules of Berger et al. while paying attention to maintain detailed balance. The authors furthermore find a hysteresis of assembly when increasing and decreasing the interaction strength and distin-

guish growth by monomers and trimers for different types of capsid structures. A further type of growth kinetics is found for the same model if capsid formation around some nanoparticle – or coiled single-stranded RNA – is considered [73]. There monomers possibly cluster around the nanoparticle and undergo collective alignment to build a closed configuration. Zhang et al. investigated two types of nucleation scenarios in a model specifically designed for accessing this problem [167]. A detailed variant of coarse-graining in the spirit of a Gō model [65, 145] is used in Ref. [18]. The basic idea is to define potential energy penalties for deviations from a minimum energy state. These potentials are adjusted by comparing the coarse-grained version of the model to a very short trajectory of an all-atom Molecular Dynamics simulation of the whole virus capsid, which is one of very few if not the only simulation in such detail reported so far. The study hints that the considered capsid might be unstable in absence of a core of coiled RNA.

1.5 Overview

Since the complete genome and therefore all single proteins are known at least in principle, gathering knowledge on the hierarchical level of protein complexes is the next big goal in molecular biology and systems biology. Dynamics is an essential property of protein complexes which is insufficiently understood so far. Although structure and dynamics can be very different, they still must be caused by the same underlying principles which are related to physics in the sense of, e.g., atomistic interactions and hydrodynamic properties of the single players. The main goal for this work is the combination of concepts from work on bimolecular protein-protein interaction and generic models for protein complexes. We want to develop a methodology which is general enough to observe complex dynamics over relevant time scales while keeping biological significance by a standardized adjustment of the model to particular systems. The simulation of whole protein complexes is computationally too expensive in the framework of an all-atom simulation. Therefore, we use a coarse-graining approach. Particularly, we avoid the use of complicated interaction potentials as these are computationally exhausting regarding numerical treatments. Each protein is regarded as a single particle with certain properties and interaction rules. Particularly, the association and dissociation is determined by the definition of spherical reactive patches. The overlap of two of such reactive patches serves as geometric criterion for an encounter complex in the model. Indeed, the encounter region typically has a more complex form as described, e.g., by Spaar et al. [140, 141]. However, the spherical shape is a simple approach and can be described with only one parameter – the patch radius. Thus, the coarse-grained model corresponds to the transport steps in the reaction network in Fig. 1.3. Each reaction step is modeled as a stochastic process. This can be adjusted to the considered system according to, e.g., more sophisticated simulations of the final association [3]. In virtually all considerations in this work we average over all possible initial conditions. This is reasonable as the biological equivalent of our systems cannot be prepared in a well known state in general. Therefore experimental results – if available – are always averages

over ensembles of many different initial configurations.

In chapter 2 the mathematical background of the model is described. The framework of this work are methods from statistical physics, particularly stochastic differential equations describing the time evolution of the system. First, we introduce the *Langevin equation* which reads like the equation of motion from classic Newtonian mechanics with an additional stochastic term representing thermal noise. We particularly consider the overdamped limit of the Langevin equation and motivate its validity on molecular scales in biological contexts. However, the cases in which an analytic solution of the equation is possible are rare. Hence, we will mostly rely on numerical methods. In particular, we describe a discretization approach to the Langevin equation and summarize the necessary details to treat it computationally. Although the principal idea of Green's function reaction dynamics [149] is not directly applicable to our model, we make use of a related approximation to derive a variable time discretization step approach. Furthermore, an effective electrostatic interaction is presented causing a systematic drift in the overdamped Langevin equation. Sect. 2.2 is dedicated to the introduction of the *Fokker-Planck equation* and first passage time problems, which is the class of problems that we mainly deal with in this work. By using the example of two finite sized and homogeneously reactive particles on a line the analytic procedures are explained. Additionally, we apply the numerical methods explained before to a Langevin representation of the same system. The comparison with the exact solution of the Fokker-Planck equation proves that the numerical treatment is correct. Finally, we summarize a general method to evaluate hydrodynamic properties of arbitrarily shaped objects. This is necessary to characterize single proteins from knowing their atomistic structure as well as estimating the properties of arbitrarily structured complexes of model proteins. A dumbbell consisting of two spherical particles at variable distance serves as an example to clarify the procedure and to check the approximation against exact results known from the literature.

In chapter 3 we introduce the details of our coarse-graining approach. We consider three different levels of detail. Particularly, proteins are modeled either as spherical particles, as dipolar spheres or as collection of several small beads with one dipole. All required parameters are obtained by a general, reproducible procedure. As three model systems with distinctly different properties we consider the pairs barnase:barstar, cytochrome c:cytochrome c peroxidase and p53:mdm2. Spherical reactive patches geometrically determining the encounter state in the model are placed on the model proteins according to the known experimental structures of the respective protein complexes. In the computer simulations, concentration is varied by changing box size. First passage times to encounter are recorded together with the number of unsuccessful contacts before encounter. We find that encounter frequency scales linearly with protein concentration, thus proving that our microscopic model results in a well-defined macroscopic encounter rate. The number of unsuccessful contacts before encounter increases with encounter rate and ranges from 20–9000. In a correlation analysis we show that the relation of the mean number of contacts scales linearly with the first passage time to encounter. For all three models, encounter

rates are obtained within one order of magnitude of the experimentally measured association rates. Electrostatic steering enhances association up to 50-fold. If diffusional encounter is dominant (p53:mdm2) or similarly important as electrostatic steering (barnase:barstar), then encounter rate decreases with decreasing patch radius. More detailed modeling of protein shapes decreases encounter rates by 5%–95%. Our study shows that the computational efficiency of modeling protein-protein encounter can be dramatically increased over atomistic approaches by using coarse-grained approaches if reactive patches are appropriately defined. We analyze the distribution of contact resting times and times of return to contact in the simulation and find a power-law $t^{-3/2}$ for small times. This is well reproduced by analytic calculations with respect to a simplified version of the model.

As our simulation scheme incorporates anisotropic diffusion by a 6×6 diffusion matrix we want to analyze its impact on protein encounter in detail. Chapter 4 starts with an explanation of anisotropic diffusion and why it is apparent only on a particular time and length scale. On a larger time scale there is a crossover to isotropic diffusion because the memory of a particular initial orientation is lost due to rotational diffusion. We review a calculation of this crossover for two-dimensional systems in a recent publication. This treatment is then extended to three dimensions. There, the situation is more complicated because the principal rotation axes rotate with the body fixed coordinate system. However, we derive an analytic expression for the crossover which matches corresponding simulation data. All of the calculations only apply for bodies that do not exhibit diffusional coupling between different degrees of freedom. As a simple model system for anisotropically diffusing particles we choose ellipsoids. These are advantageous in two ways. First, analytic results for the principal diffusion coefficients are known from the literature and there is no diffusional coupling for ellipsoids. Furthermore, a rigorous mathematical criterion exists to check whether two ellipsoids overlap. This can be used to conveniently implement the excluded volume condition. To investigate the impact of anisotropy we consider the encounter of pairs of ellipsoids with varying aspect ratios in a periodic boundary simulation box. The encounter is again defined by the overlap of spherical reactive patches placed at distinct positions at the surface of the ellipsoids. Particularly, we consider patches located at the apex of one of the two distinct semi-axes. Therefore, three different combinations of patch locations exist for a pair of ellipsoids. We pay attention to the altered overall mobility for differing aspect ratios. Furthermore, by a generic measure of the steric accessibility of the reactive patches we show that the main contribution for the altered encounter rates at different ratios is due to excluded volume effects. Finally, we compare the data with a different set of simulations where the steric form of the ellipsoids is considered, but not the hydrodynamic anisotropy. We only find a significant deviation for prolates with a patch located on the symmetry axis. This proves that steric effects rather than hydrodynamics dominate anisotropic encounter.

The methodology we use in this thesis is suited to model large complexes. Chapter 5 is about the dynamics beyond bimolecular reactions. We start with the study of three homogeneously binding particles on a line with periodic boundary conditions in analogy to Sect. 2.2. We find that the mean first passage time to

the first encounter of two of the particles cannot be calculated with a differential equation of Poisson type with respect to two relative distances. However, it is possible to derive a similar equation containing a term of mixed derivatives to solve the problem. The solution has a similar form as in the case of two particles and matches corresponding simulation data. Due to the higher complexity the calculation of first passage time distributions is not possible. We find that the formation of the first bond biases the second binding event, which forms a single cluster of all three particles. Therefore, a simple ansatz of two conceptionally equal processes does not lead to the results obtained by corresponding computer simulations. Such biases will also occur in more complex situations. In three dimensions, the assembly of model proteins requires a set of rules regarding the relative alignment after binding. Particularly, in our approach we do not consider slow dynamic changes of the alignment, but rather choose an instantaneous switch of the two proteins into the predefined relative position. This also affects proteins which are already attached to the binding partners. These simulation rules are applied to a three-particle cluster in three dimensions. We try to establish an assembly network in the spirit of Fig. 1.3. However, we find that especially the transitions in the network representing transport processes are not Poisson-like. This particularly shows that it is not allowed in general to describe assembly processes by a respective Master equation. Finally, large scale simulations of self-assembly processes of virus capsids are performed. We find that capsid assembly is most efficient at an intermediate range of the unbinding rate of the monomers. The chapter is closed by an outlook regarding particular challenges for the treatment of large complexes.

Chapter 2

Model and methods

In this chapter we introduce the theoretical framework of our model. In the first part, two types of stochastic differential equations are explained. On the one hand the Langevin equation is discussed particularly in the overdamped limit. With regard to the numerical solution a discretized approach is described and some details of the numerics are summarized. The Fokker-Planck equation can be used to analytically study first passage time problems. Both are reviewed in the second section. We present analytic solutions of the association dynamics of homogeneously binding particles in two well-defined topologies. Our solutions deal as check for the simulation technique described in the first part. Finally, we summarize a sophisticated method to evaluate hydrodynamic properties of arbitrarily shaped objects, which we use in this work. The results of this approximate method for the simple geometry of the dumbbell are compared to exact results known from the literature and serve as a measure for its reliability.

2.1 Langevin equation approach

Theoretical basis

A standard approach for the description of the thermally driven Brownian motion of particles in the framework of statistical mechanics is the Langevin equation [78]. Provided that thermal forces are spatially homogeneous a general form is:

$$m\,\frac{\partial}{\partial t}\mathbf{v}_t = \mathbf{F}_t - \zeta\mathbf{v}_t + \mathbf{K}_t\ . \tag{2.1}$$

In Eq. (2.1) $\mathbf{v}_t$ denotes the speed of a particle in solution, m is its mass, ζ is the friction matrix which will be explained in more detail later, $\mathbf{F}_t$ is the sum of all forces directly acting on the particle while $\mathbf{K}_t$ denotes a stochastic force due to collisions with molecules from the solution. The subscript t is meant to explicitely point out the time dependency of the quantities. derivative. Without the noise term $\mathbf{K}_t$, Eq. (2.1) is simply Newton's second law. $\mathbf{K}_t$ can be considered as uncorrelated Gaussian noise with zero mean, also called *white noise*. If $\mathbf{K}_t$ does not depend on position $\mathbf{x}$, then the thermal noise is purely *additive*, as it is only

present in an independent term in the above equation. Inhomogeneous, i.e., position dependent thermal forces $\mathbf{K}_t = \mathbf{K}_t(\mathbf{x})$ cause an additional drift term and lead to so called *multiplicative noise*. This is more difficult to handle as it requires an additional definition of the interpretation of the noise term (for details see Refs. [78, 148]). An example is the motion of a particle above a wall, where the friction changes according to the height [85, 86]. If m/ζ is much smaller than the characteristic time scale of $\mathbf{F}_t$ and $\mathbf{K}_t$, then the effects of inertia are negligible due to the damping term $-\zeta\mathbf{v}_t$. Typically, this assumption is valid on molecular scales in biological systems, where the mass of particles is small and the friction due to the high density of water molecules is high. That is, in the context relevant for this work the equation can be viewed in the *overdamped* limit, where $m\,\partial/\partial t(\mathbf{v}_t) \to 0$ and Eq. (2.1) can thus be written as:

$$\mathbf{v}_t = \frac{\partial}{\partial t}\mathbf{x}_t = \zeta^{-1}\left(\mathbf{F}_t + \mathbf{K}_t\right) \ . \tag{2.2}$$

The inverse of the friction matrix is also called the mobility matrix $\zeta^{-1} = \mathbb{M}$. In thermal equilibrium the mobility matrix is directly connected to the diffusion matrix via the Einstein-Smoluchowski relation $\mathbb{D} = k_B T_a \mathbb{M}$ [48, 138], where $k_B T_a$ is the thermal energy at ambient temperature T_a and k_B is the Boltzmann constant. Under non-equilibrium conditions a more general fluctuation dissipation theorem applies [142]. Eq. (2.2) is also valid in a generalized coordinate space where the vectors are six-dimensional and do not only account for the translational motion of the particle but also for its rotation. In the following, let $\mathbf{X}_t$ be a six-dimensional vector describing position and orientation of a particle at time t. Furthermore, $\mathbf{F}_t$ and $\mathbf{K}_t$ shall be generalized six-dimensional combinations of the respective forces and torques. The latter three components of $\partial/\partial t(\mathbf{X}_t)$ thus give a vector $\boldsymbol{\omega}_t$ which determines the rotation of the particle. $\boldsymbol{\omega}_t$ points into the direction of the rotation axis and $|\boldsymbol{\omega}_t|$ defines the angle of rotation. After these considerations, Eq. (2.2) can be written as:

$$\frac{\partial}{\partial t}\mathbf{X}_t = \mathbb{M}\mathbf{F}_t + \mathbf{g}_t \ . \tag{2.3}$$

In this notation the noise term is $\mathbf{g}_t = \mathbb{M}\mathbf{K}_t$. As $\mathbf{K}_t$ is Gaussian white noise, the distribution of $\mathbf{g}_t$ is fully determined by its first two moments because all higher moments vanish:

$$\langle \mathbf{g}_t \rangle = 0 \ , \qquad \langle \mathbf{g}_t \mathbf{g}_{t'} \rangle = 2k_B T_a \mathbb{M}\delta(t-t') = 2\mathbb{D}\delta(t-t') \ . \tag{2.4}$$

Formally one can conclude from Eq. (2.4) that $\mathbf{g}^2 \sim \mathbb{M}$ and thus $\mathbf{g} \sim \sqrt{\mathbb{M}}$. Bearing in mind this finding Eq. (2.3) shows that while the Brownian contribution to motion is proportional to $\sqrt{\mathbb{M}}$, the direct impact of force and torque increases linearly with $\mathbb{M}$. That is, in the regime of high friction and thus low mobility, the stochastic Brownian motion of the particles can get more important than the impact of $\mathbf{F}_t$. An example with a rough estimation of this effect by typical numbers will be given in the discussion of the effective electrostatic interaction between macromolecules at the end of this section. We now explain a rigorous way to handle the integration of Eq. (2.3).

Algorithmic details

One approach to the solution of Eq. (2.3) is to consider a discretized form according to the Euler algorithm:

$$\mathbf{X}(t+\Delta t) = \mathbf{X}(t) + \mathbb{M}\mathbf{F}(t)\Delta t + \mathbf{g}(\Delta t) + \mathcal{O}(\Delta t^2) \; . \tag{2.5}$$

Note that $\mathbf{g}(\Delta t)$ is defined similar to $\mathbf{g}_t$ in Eq. (2.4) but $\delta(t-t')$ goes over to a factor Δt:

$$\langle \mathbf{g}(\Delta t) \rangle = 0 \; , \qquad \langle \mathbf{g}(\Delta t)\mathbf{g}(\Delta t) \rangle = 2\mathbb{D}\Delta t \; . \tag{2.6}$$

Thus it explicitely depends on the size of the time step $\mathbf{g}(\Delta t) \sim \sqrt{\Delta t}$. The definition in Eq. (2.6) is not straightforward to implement, but Ref. [53] gives a recipe to compute random vectors $\mathbf{g}$ whose components reproduce the correct distribution. A more detailed explanation is given in Ref. [86], where the same approach has been used to model cell adhesion in shear flow via reactive receptor patches. The key is to compute the Cholesky decomposition $\mathbb{B}$ of $\mathbb{M}$ which satisfies $\mathbb{M} = \mathbb{B}\mathbb{B}^\dagger$, where $\mathbb{B}^\dagger$ is the transposed of matrix $\mathbb{B}$. Now a vector of six normal deviate random numbers $\mathbf{x}$ is computed, whose components satisfy $\langle x_i \rangle = 0$, $\langle x_i x_j \rangle = 2k_B T_a \Delta t \, \delta_{ij}$. $\mathbf{g}(\Delta t)$ is then given by:

$$\mathbf{g}(\Delta t) = \mathbb{B}\mathbf{x} \; . \tag{2.7}$$

Note that as Eq. (2.5) is only correct up to order $\mathcal{O}(\Delta t^2)$, actually only the first two moments of $\mathbf{g}(\Delta t)$ have to be reproduced correctly [78]. Thus, it is not necessary to chose $\mathbf{x}$ from a normal distribution as it would be possible, e.g., by the Box-Müller method [28]. In this special case it is equivalent and computationally less costly to calculate equally distributed random numbers $x_i = \sqrt{24 k_B T_a \Delta t}(r_i - 0.5)$, where r_i denotes a uniformly distributed random number $0 \leq r_i < 1$. The typical orders of magnitude for translational diffusion coefficients of proteins are $D = 10^{-6}\mathrm{cm}^2\mathrm{s}^{-1}$ and the radii are in the nanometer range. Therefore, a reasonable choice for the time step is $\Delta t = 1\mathrm{ps}$, as this leads to a root mean square deviation due to Brownian motion of $\sqrt{D\Delta t} = 0.01\mathrm{nm}$, which is well below the size of the proteins.

The mobility matrix of a particle must be defined in a particle-fixed coordinate system to be independent of its orientation. Thus, the displacement has to be calculated in terms of particle-fixed coordinates, and then transformed to the laboratory coordinate space. Particularly, this transformation means a rotation $\mathbb{R}$ according to the orientation of the particle. Special attention has to be payed regarding the force $\mathbf{F}$, which is typically calculated in the global frame of reference and hence has to be transformed to particle space first, before Eq. (2.5) can be evaluated. This back-transformation is achieved by applying $\mathbb{R}^{-1}$ to $\mathbf{F}$. Since rotation matrices $\mathbb{R}$ simply consist of a list of orthonormal vectors they are orthogonal so that their inverse is equal to the transposed matrix $\mathbb{R}^{-1} = \mathbb{R}^\dagger$. After these considerations, Eq. (2.5) should be rewritten to:

$$\mathbf{X}(t+\Delta t) = \mathbf{X}(t) + \mathbb{R}\left[\mathbb{M}\left(\mathbb{R}^\dagger \mathbf{F}(t)\right)\Delta t + \mathbf{g}(\Delta t)\right] + \mathcal{O}(\Delta t^2) \; . \tag{2.8}$$

Note that as $\mathbf{F}$ and g are six-dimensional and contain information about torque and rotation, Eq. (2.8) is only formally correct, because $\mathbb{R}$ acts on both the translational and rotational parts of the respective vectors.

Variable time steps

In many biological cases, the concentration of the reacting proteins is rather low. The association rate typically scales linearly with the concentration and thus low concentrations will lead to large mean association times. Thus, a lot of computer time will be spent on situations, where all proteins are purely diffusing without much interaction. For this reason we use a variable time step in most of our simulations. Van Zon and ten Wolde suggested a method to control the next collision when they introduced their Green's function reaction dynamics (GFRD) [150]. In contrast to our work, however, this method is based on isotropic diffusion. Generalizing the GFRD to anisotropic diffusion is very difficult and out of the scope of our work. We therefore used the following scheme. We first note that in GFRD each time step is chosen such that it includes the next reaction. In our case, we also want to investigate the stochastic dynamics before the next encounter event takes place. Thus a large time step is not chosen to include the next encounter, but to bring the system to such a configuration that encounter becomes more likely. This step can be well represented by isotropic diffusion with an overall diffusion constant $D = \mathbb{D}_{11} + \mathbb{D}_{22} + \mathbb{D}_{33}/3$ following from the translational part of the anisotropic diffusion matrix. For an isotropic random walk, the displacement probability is given by a Gaussian distribution with spherical symmetry. Particularly, the mean distance is $r = 0$ and large displacements are exponentially suppressed according to the standard deviation $\sqrt{6D\Delta t}$. Therefore, a displacement of the magnitude $\Delta r^H_{\max} \geq H\sqrt{6D\Delta t}$ is H standard deviations apart from the average and thus occurs with a probability of $\sim e^{-H}$. We define the effective particle distance as the distance of the surfaces $r^{\text{eff}}_{ij} = |\mathbf{r}_i - \mathbf{r}_j| - R_i - R_j$, where R_i determines the maximal steric interaction radius of particle i. By setting $\Delta r^H_{\max} = \min\{r^{\text{eff}}_{ij}\}$ the smallest effective particle distance in the system one can estimate a reasonable time step for which a collision is hardly probable. Van Zon and ten Wolde found that the choice $H = 3$ provides good results. Besides the small probability of accidentally drawing a displacement that is large enough, two particles can only collide if they move *towards* each other. This further lowers the probability of sampling erroneous collisions as the direction of motion is completely random. When the particles reach close proximity $\min\{r^{\text{eff}}_{ij}\} \to 0$ the estimated step size Δt vanishes and thus the simulation would be infinitely slowed down. Therefore, there has to be a lower boundary for the time step $\Delta t_{\min}$, for which we generally choose $\Delta t_{\min} = 1\text{ps}$ as reasoned earlier. Thus, the adapted time step is given by:

$$\Delta t_{\text{ad}} = \min\left\{\frac{H^2}{6D}\left(\min\{r^{\text{eff}}_{ij}\}\right)^2; \Delta t_{\min}\right\} . \tag{2.9}$$

The 6×6 diffusion matrix $\mathbb{D}$ represents anisotropic diffusion. For large times, anisotropic diffusion effectively crosses over into isotropic diffusion because the information about the initial orientation gets lost after a certain relaxation time

due to the rotational diffusion [75]. In general, translational and rotational diffusion are coupled so that large time steps cannot be used. This problem will be treated in detail in chapter 4. However, for particular protein systems as studied in chapter 3, e.g., one finds that the diffusive coupling is a rather small effect. In particular, the major entries in the diffusion matrix of smaller proteins scaled by different powers of the typical Stokes radius $R \sim 10^{-7}\mathrm{cm}$ to make the dimensions comparable are $D_{\mathrm{tt}}/R^2 \sim 10^8\mathrm{s}^{-1}$, $D_{\mathrm{rr}} \sim 10^7\mathrm{s}^{-1}$, $D_{\mathrm{tr}}/R \sim 10^5\mathrm{s}^{-1}$. Therefore, the effect of diffusive coupling is 10^{-2} and 10^{-3} smaller than rotational and translational diffusion, respectively. Finally, the typical time scale at which the cross-over is expected can be calculated to be $1/(6D_{\mathrm{rr}}) \approx 10\mathrm{ns}$ (see chapter 4). This is 10^4 times as large as the minimal time step according to the previous considerations. Time steps of this magnitude were rarely found in the simulations, so that for most of the steps, the anisotropy is well preserved. Therefore, we can safely neglect changes in the anisotropy of the effective mobility matrix when using larger time steps.

Pair interactions

Especially at small separations a multitude of complex effects play a role in protein-protein interactions, which have to be considered at the atomisitc level (see Sect. 1.2). However, as the encounter complex typically means inter-protein separations of several angstroms, it still can be well described while neglecting some of the details. In this work we focus on the two most important types of interactions. One is simply arising by an excluded volume constraint. The other is the electrostatic interaction of the model proteins.

The excluded volume effect is integrated into the numerical integration scheme of the Langevin equation by a Monte Carlo like approach. In each step of a simulation, a displacement vector $\Delta\mathbf{X}(t)$ is drawn for each particle as described above. If this global displacement leads to any violation of the hardcore repulsion, all suggested displacements are rejected and new set of $\Delta\mathbf{X}(t)$ is calculated. This procedure continues, until an update of all positions and orientations is found which does not lead to any overlap. In this way, the constraint according to the excluded volume effect is included in the stochastic motion. In general one would have to solve the scattering problem. Note, however, that for proteins it is not clear how to proceed, thus less rigorous approach seems appropriate. One would expect that our procedure leads to errors of order Δt if two particles are in close proximity of order $\sqrt{D\Delta t}$. However, it has been shown for a different system [115] that in practice the deviation from the expected behavior is very small and thus the approach is reasonable.

Coulombic interactions are known to play a prominent role in protein association (see e.g. Ref. [99]). They are caused by the sum of all atomistic charges and their higher electrostatic moments. With a typical number of 10^4 atoms in a protein, the exact evaluation of all $\sim 10^4 \times 10^4$ pair interactions needs a lot of computer time, while many of them might be negligible due to the electrostatic screening. Following Refs. [51, 71] the dipolar sphere model (DSM) can be used to effectively model the interaction according to the total monopole and dipole

moments of all charges in each model protein:

$$\hat{q} = \sum_i q^i \ , \qquad \hat{\mathbf{p}} = \sum_i q^i \mathbf{r}(q^i) \ , \tag{2.10}$$

where i counts over the atomistic charges q^i and $\mathbf{r}(q^i)$ denotes the position of charge i relative to the center of mass. In the DSM $\hat{q}$ is represented by a single charge $q_1 = \hat{q}$ located at the center of mass carrying the sum of all charges of the protein. $\hat{\mathbf{p}}$ is represented by two opposing charges $q_{2/3}$ at some distance r_p from the center of mass. In the original work by Eltis et al. [51] the authors choose r_p such that $q_{2/3}$ are located at a distance of 1.5Å beneath the surface of the respective model protein. We checked different definitions and found that the particular choice of r_p does not crucially change the simulation results. Therefore, we placed $q_{2/3}$ at 4.0Å beneath the surface of the model proteins of our simulations including the DSM unless otherwise stated. The magnitude and location of the model charges is:

$$q_{2/3} = \pm \frac{1}{2r_p} |\hat{\mathbf{p}}| \ , \qquad \mathbf{r}(q_{2/3}) = \frac{1}{2q_{2/3}} \hat{\mathbf{p}} \ . \tag{2.11}$$

The factor 2 in the denominator results from the fact that $q_{2/3}$ both carry half of the necessary charge to reproduce:

$$q_2 \mathbf{r}(q_2) + q_3 \mathbf{r}(q_3) = \left(\frac{q_2}{2q_2} + \frac{q_3}{2q_3} \right) \hat{\mathbf{p}} = \hat{\mathbf{p}} \ , \tag{2.12}$$

$$\text{while} \quad |\mathbf{r}(q_{2/3})| = \frac{1}{2\frac{1}{2r_p}|\hat{\mathbf{p}}|} |\hat{\mathbf{p}}| = r_p \ . \tag{2.13}$$

Taking into account the Debye screening function due to the presence of counter ions in solution, the electrostatic interaction energy between two charges $q_{i/j}$ at positions $\mathbf{r}_{i/j}$ respectively with distance $r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_j - \mathbf{r}_i|$ is:

$$W_{ij} = \frac{1}{4\pi\varepsilon_0\varepsilon_r} q_i q_j \frac{e^{-\kappa(r_{ij} - B_{ij})}}{(1 + \kappa B_{ij}) r_{ij}} \ . \tag{2.14}$$

Here, $\kappa = l_D^{-1}$ is the inverse Debye screening length, which typically has a value of $\approx 1\text{nm}$ under physiological conditions. B_{ij} is an approximate correction to the screening for charges which are placed in an object like a protein, since within the protein, there are no free counter ions, so for the screening only the distance between two charges outside the containing proteins must be considered. Taking $b_{i/j}$ as the closest distance of $q_{i/j}$ from the surface of the surrounding protein, it is approximately given by $B_{ij} = b_i + b_j$. Assuming a roughly spherical form of the protein with radius R this length would be given by $b_1 = R$ and $b_{2/3} = R - r_p$. This potential leads to a force of charge q_j on q_i:

$$\begin{aligned} \mathbf{F}_{ij} &= -\nabla_{\mathbf{r}_i} W_{ij} = -\frac{\partial W_{ij}}{\partial r_{ij}} \cdot (\nabla_{\mathbf{r}_i} r_{ij}) \\ &= -\frac{1}{4\pi\varepsilon_0\varepsilon_r} q_i q_j \frac{e^{-\kappa(r_{ij} - B_{ij})}(1 + \kappa r_{ij})}{(1 + \kappa B_{ij}) r_{ij}^2} \frac{\mathbf{r}_{ij}}{r_{ij}} \ . \end{aligned} \tag{2.15}$$

Two model particles m and n feel the sum of the Coulomb forces $\mathbf{F}_{ij}$ between all pairs of the three complementary charges mimicking the monopolar and dipolar interactions. That is, the full force between particle m and n is $\mathbf{F}_{mn} = \sum_{i=1}^{3}\sum_{j=1}^{3}\mathbf{F}_{ij}$, where i/j count over the charges of m/n respectively.

The impact of the force on a particle in the Langevin equation is weighted with the mobility matrix, which is the inverse of the friction matrix $\mathbb{M} = \zeta^{-1} = \mathbb{D}/(k_B T_a)$. Charges q are always given in multiples of the elementary charge e. From Eq. (2.5) and Eq. (2.6) we see that the contribution of the Brownian term $\mathbf{g}(\Delta t)$ and the force $\mathbb{M}\mathbf{F}\Delta t$ are given in dimensions of length, i.e., as direct summand to $\Delta\mathbf{X}(\Delta t)$. The fixed prefactors for the displacement due to the force term give a length of $e^2/(4\pi\varepsilon_0 k_B T_a) = 5.7031 \cdot 10^{-8}\mathrm{m}$ at $T_a = 293\mathrm{K}$. The remaining factors then give a dimensionless quantity proportional to $\sim D\Delta t/r_{ij}^2$. As a rough estimate for the the impact of the force, we compare the typical displacement due to the Brownian motion and due to the force between two elementary charges at a distance $r_{ij} = 1\mathrm{nm}$ and $r_{ij} = 4\mathrm{nm}$, neglecting B_{ij}. A typical size of a protein is in the range of $R = 1\mathrm{nm}$, which gives a diffusion constant of the order of $D = 10^{-6}\mathrm{cm}^2/\mathrm{s}$. Considering a time step of $\Delta t = 1\mathrm{ps}$, $D|\mathbf{F}(1\mathrm{nm})|\Delta t/k_B T_a \sim 10^{-14}\mathrm{m}$ and $D|\mathbf{F}(4\mathrm{nm})|\Delta t/k_B T_a \sim 10^{-16}\mathrm{m}$, while the typical step length due to the Brownian motion is $\sqrt{D\Delta t} \sim 10^{-11}\mathrm{m}$. This shows that the magnitude of electrostatic interactions at distances of $1\mathrm{nm}$ (and more so at $4\mathrm{nm}$)is much smaller than the thermal energy. However, we expect that the *systematic* drift, while small, will still lead to an altered encounter behavior.

2.2 First passage time problems

2.2.1 The Fokker-Planck equation

A common class of problems in the framework of stochastic equations like the Langevin equation are *first passage time* (FPT) problems. It refers to the time to the first match of the stochastically evolving system with a certain boundary condition – the *passage* – having started at some initial configuration. The stochastic dynamics of the system results in a stochastic component in the process of the first passage. It will be different for every particular trajectory. Therefore, the first passage time itself is a random variable. One interesting quantity with respect to the first passage time is its average value, the mean first passage time (MFPT). The full FPT distribution is more meaningful but also typically much harder to calculate analytically. However, both can be possibly obtained via simulation. The mathematical solution of the first passage time problem usually incorporates the solution of the stochastic problem with a certain set of boundary conditions. In the following, we will particularly consider the *Fokker-Planck equation* (FPE), which can be shown to be equivalent to the Langevin equation. Let $p(\mathbf{x}, t)$ be the probability density of the system. Typically, the initial condition for such a process is to let the system start at some particular position $\mathbf{x}_0$ at time $t = 0$. Strictly, p then denotes a transition probability from state $(\mathbf{x}_0, 0)$ to $(\mathbf{x}, t)$ and is commonly written as $p(\mathbf{x}, t|\mathbf{x}_0, 0)$. However, in the following we presume $(\mathbf{x}_0, 0)$ as initial condition and

simply write $p(\mathbf{x}, t)$. The time evolution is then given by the continuity equation:

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = -\nabla \mathbf{j}(\mathbf{x}, t) \ , \tag{2.16}$$

where $\mathbf{j}$ is the probability flux:

$$\mathbf{j}(\mathbf{x}, t) = \left(\mathbf{A}(\mathbf{x}, t) - \frac{1}{2} \nabla D(\mathbf{x}, t) \right) p(\mathbf{x}, t) \ . \tag{2.17}$$

Here, $\mathbf{A}$ is the *drift*, e.g. caused by the action of a potential U:

$$\mathbf{A}(\mathbf{x}, t) = -\nabla U(\mathbf{x}, t) \ . \tag{2.18}$$

The *diffusion* term $\nabla D(\mathbf{x}, t)$ contains the stochasticity in terms of a dynamics driven by temperature. In general, the scalar $D(\mathbf{x}, t)$ explicitely depends on time and space. However, in the cases considered in this work, there are no such dependencies, i.e., $D(\mathbf{x}, t) \equiv D$. The factor $1/2$ in Eq. (2.17) is a convention [78]. With constant D we can rewrite Eq. (2.17) in form of a heat equation:

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = \left(\frac{D}{2} \triangle - \nabla \mathbf{A}(\mathbf{x}, t) \right) p(\mathbf{x}, t) \ . \tag{2.19}$$

Note that the ∇ operator in the drift term acts on both $\mathbf{A}$ and p. There is also the possibility to formulate the FPE for some fixed end configuration with respect to the initial conditions, which gives the so called *adjoint* FPE:

$$\frac{\partial}{\partial t_0} p(\mathbf{x}, t | \mathbf{x}_0, t_0) = \left(-\frac{D}{2} \triangle_{\mathbf{x}_0} - \mathbf{A}(\mathbf{x}_0, t_0) \nabla_{\mathbf{x}_0} \right) p(\mathbf{x}, t | \mathbf{x}_0, t_0) \ . \tag{2.20}$$

Note that now the operators $\triangle_{\mathbf{x}_0}$ and $\nabla_{\mathbf{x}_0}$ act *only* on p.

Mathematically, the passage is described by an *absorbing boundary* in the system with respect to the probability density $p(\mathbf{x}, t)$, i.e., the latter is set to zero at the boundary $\partial\Omega$:

$$p(\mathbf{x}, t)\big|_{\mathbf{x} \in \partial\Omega} \equiv 0 \ . \tag{2.21}$$

A more general treatment is the so called *radiation boundary condition* which is incorporated by correlating the outward probability flux $\mathbf{j}(\mathbf{x}, t)$ to some finite absorption rate k times the probability density at the boundary $p(\mathbf{x}, t)$:

$$\mathbf{n}(\mathbf{x}) \cdot \mathbf{j}(\mathbf{x}, t) = k p(\mathbf{x}, t) \ , \qquad \mathbf{x} \in \partial\Omega \ . \tag{2.22}$$

Here, $\mathbf{n}(\mathbf{x})$ denotes the normal vector of the boundary $\partial\Omega$. More generally, k could also contain a time or space dependence $k = k(\mathbf{x}, t)$. With this one can, e.g., model a reaction of two particles happening only at a certain finite rate constant k (see Ref. [144], e.g.). Particularly, for large rates $k \to \infty$ one ends at the absorbing boundary condition $p \sim \mathbf{j}/k \to 0$ as given in Eq. (2.21). In the limit of very small rates $k \to 0$ the boundary is getting impermeable and the flux vanishes. This is called a *reflecting boundary*:

$$\mathbf{n}(\mathbf{x}) \cdot \mathbf{j}(\mathbf{x}, t) \equiv 0 \ , \qquad \mathbf{x} \in \partial\Omega \ . \tag{2.23}$$

Having solved Eq. (2.19) with the appropriate boundary and initial conditions to reflect the particular problem, the FPT distribution can then be calculated from p. The total probability G of the particle to have not yet been absorbed at some time t is the integral of p over the whole accessible region Ω:

$$G(t) = \int_{\Omega} \mathrm{d}\mathbf{x}'\, p(\mathbf{x}', t) \; . \tag{2.24}$$

In each infinitesimal time interval G will be lowered by some amount, which reflects the part of the probability density captured by the absorbing boundary. The larger this part the higher is the probability of being captured in the particular time interval. That is, the probability of being absorbed at time t can be expressed as the negative time derivative of $G(t)$:

$$f(t) = -\dot{G}(t) = -\int_{\Omega} \mathrm{d}\mathbf{x}'\, \dot{p}(\mathbf{x}', t) \; . \tag{2.25}$$

The mean first passage time (MFPT) is equivalent to the first moment of the FPT distribution:

$$T_1(\mathbf{x}_0) = T(\mathbf{x}_0) = \int_0^{\infty} \mathrm{d}t'\, t' f(t') \; . \tag{2.26}$$

Note that in our formulation T_1 is a scalar without any *explicit* dependence on the initial configuration $\mathbf{x}_0$. The dependence is part of the initial value problem that has to be solved to find $f(t)$. In Ref. [78] a convenient way is described to express higher moments. The key finding is that the n^{th} moment T_n can be written as:

$$T_n = \int_0^{\infty} \mathrm{d}t'\, t'^n f(t') = -\int_0^{\infty} \mathrm{d}t'\, t'^n \dot{G}(t') = n \int_0^{\infty} \mathrm{d}t'\, t'^{n-1} G(t') \; , \tag{2.27}$$

for $n \geq 1$ and $T_0 \equiv 1$. Combining Eq. (2.20) and Eq. (2.24) leads to:

$$\frac{\partial}{\partial t} G(t|\mathbf{x}_0, t_0) = \left(\mathbf{A}(\mathbf{x}_0, t_0) \nabla_{\mathbf{x}_0} + \frac{D}{2} \triangle_{\mathbf{x}_0} \right) G(t|\mathbf{x}_0, t_0) \; . \tag{2.28}$$

Assuming a drift constant in time $\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x})$ and multiplying both sides of the equation with t^{n-1} one can integrate Eq. (2.28) with respect to t and use Eq. (2.27) to obtain:

$$\left(\mathbf{A}(\mathbf{x}_0) \nabla_{\mathbf{x}_0} + \frac{D}{2} \triangle_{\mathbf{x}_0} \right) T_n(x_0) = -n T_{n-1}(x_0) \; , \tag{2.29}$$

for $n \geq 1$ and with $T_0 \equiv 1$. Given this result, one then can proceed to higher moments in a recursive manner. Particularly, without drift $A(x_0) \equiv 0$ the MFPT is the solution of the Poisson type equation:

$$\triangle_{\mathbf{x}_0} T(\mathbf{x}_0) = -\frac{2}{D} \; . \tag{2.30}$$

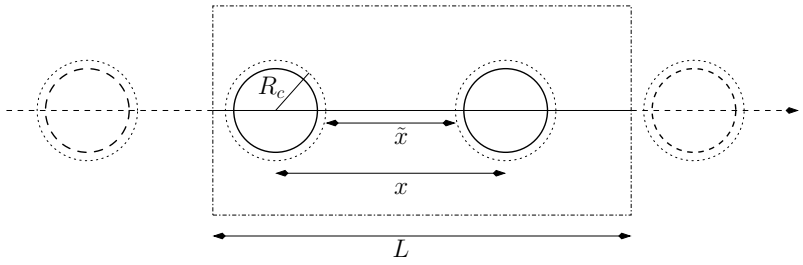


Figure 2.1: Simple model system with two finite sized particles moving on a one-dimensional track. Periodic boundary conditions are hinted by the box and the two virtual images of the particles outside the boundaries are given with dashed lines. The dotted lines denote the binding radius.

2.2.2 One-dimensional diffusion with periodic boundary conditions

As a first simple test of the implementation of our model, we want to study a system of two equal-sized, homogeneously binding particles, whose movement is confined to one dimension with periodic boundary conditions (see Fig. 2.1). We consider the motion as free, i.e., no drift is taken into account. As throughout this work we neglect hydrodynamic interactions between the particles but take the mobility into account as derived from the Stokes equation, which applies at the considered scales. Particularly, the mobility of a sphere according to Stokes' law is $M_t = 1/(6\pi\eta_0 R)$, where R is the radius of the sphere and η_0 is the viscosity of the fluid. As we work with dimensionless quantities here, we identify diffusion and mobility $D_t = M_t$, i.e., we choose the thermal energy such that it satisfies $2k_B T_a = 1$ (compare Eq. (2.4)). In our example it is sufficient to consider the relative position x of the two spheres, which undergoes a random walk with a diffusion constant $D = 2D_t$. Denoting the binding radius of each particle with R_c, the two spheres bind if $x = x_r = 2R_c$. Due to the periodic boundary conditions, this is equivalent to the situation, where $x = x_l = L - 2R_c$ with L being the length of periodicity. We consider the binding event as the passage of the system.

According to the one-dimensional representation of Eq. (2.30), the MFPT T is the solution of the ordinary differential equation:

$$\frac{\partial^2}{\partial x_0^2} T(x_0) = -\frac{2}{D} \,, \tag{2.31}$$

where x_0 denotes the initial relative position of the two particles. For simplicity, we shift $x \to \tilde{x} = x - 2R_c$ in the following, which gives $\tilde{x}_r = 0$ and $\tilde{x}_l = L - 4R_c = \tilde{L}$. At the boundaries, it holds $T(\tilde{x}_r) = T(\tilde{x}_l) = 0$. This leads to a symmetric solution:

$$T(\tilde{x}_0) = \frac{1}{D}\left(\tilde{L} - \tilde{x}_0\right)\tilde{x}_0 \,. \tag{2.32}$$

To calculate the FPT distribution we first ask for the probability density, which is the solution of:

$$\frac{\partial p(x,t)}{\partial t} = \frac{D}{2}\frac{\partial^2 p(x,t)}{\partial x^2} \,. \tag{2.33}$$

Eq. (2.33) has the form of the heat equation, which can have solutions of the form [113]:

$$p(x,t) = A\exp\left(-\mu^2 Dt/2\right)\cos\left(\mu x + B\right) + C\ , \tag{2.34}$$

where A, B, C and μ are constants determined by the initial conditions. Using an infinite sum of solutions of type (2.34) makes it possible to have a Fourier series approach to any kind of initial conditions. First, we apply the boundary conditions. From $p(\tilde{x}_r,t) = p(\tilde{x}_l,t) = 0\,\forall\, t$ it follows $B = \pi/2$ and $\mu = \pi n/\tilde{L}$ with $n \in \mathbb{N}$. Additionally assuming $C = 0$ the complete solution is (compare [118]):

$$p(\tilde{x},t) = \sum_{n=1}^{\infty} A_n \exp\left(-\frac{D}{2}\left(\frac{\pi n}{\tilde{L}}\right)^2 t\right)\sin\left(\frac{\pi n}{\tilde{L}}\tilde{x}\right)\ . \tag{2.35}$$

The standard initial condition is to start at a certain location $0 \le \tilde{x}_0 \le \tilde{L}$ at time $t = 0$, which means $p(\tilde{x},0|\tilde{x}_0,0) = \delta(\tilde{x}-\tilde{x}_0)$. According to e.g. Ref. [17], the Fourier coefficients A_n can be calculated via:

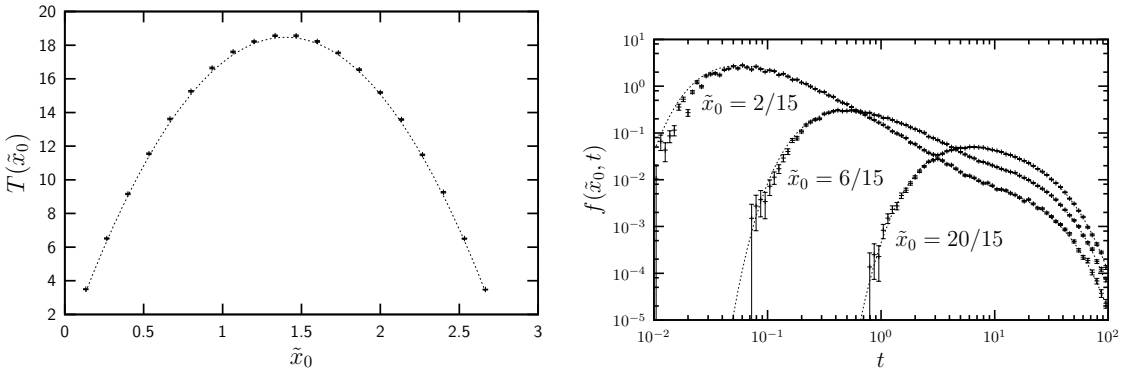
$$A_n = 2\frac{1}{\tilde{L}}\int_0^{2\tilde{L}} \mathrm{d}x'\,\delta(x'-\tilde{x}_0)\sin\left(\frac{\pi n}{\tilde{L}}x'\right) = \frac{2}{\tilde{L}}\sin\left(\frac{\pi n}{\tilde{L}}\tilde{x}_0\right)\ . \tag{2.36}$$

The first passage time distribution $f(\tilde{x}_0,t)$ is the negative time derivative of the probability that the system is still located between the boundaries $0 < \tilde{x} < \tilde{L}$ at time t:

$$\begin{aligned} f(\tilde{x}_0,t) &= -\frac{\partial}{\partial t}\int_0^{\tilde{L}} \mathrm{d}x'\,p(\tilde{x}',t|\tilde{x}_0,0) \\ &= -\sum_{n=1}^{\infty}\frac{2}{\tilde{L}}\sin\left(\frac{\pi n}{\tilde{L}}\tilde{x}_0\right)\left[\frac{\partial}{\partial t}\exp\left(-\frac{D}{2}\left(\frac{\pi n}{\tilde{L}}\right)^2 t\right)\right]\left[\int_0^{\tilde{L}}\mathrm{d}x'\,\sin\left(\frac{\pi n}{\tilde{L}}\tilde{x}'\right)\right] \\ &= \sum_{n=1}^{\infty}\frac{n\pi D}{\tilde{L}^2}\left(1-\cos\left(n\pi\right)\right)\sin\left(\frac{\pi n}{\tilde{L}}\tilde{x}_0\right)\exp\left(-\frac{D}{2}\left(\frac{\pi n}{\tilde{L}}\right)^2 t\right) \qquad (2.37) \\ &= \sum_{n=0}^{\infty}\frac{2(2n+1)\pi D}{\tilde{L}^2}\sin\left(\frac{\pi(2n+1)}{\tilde{L}}\tilde{x}_0\right)\exp\left(-\frac{D}{2}\left(\frac{\pi(2n+1)}{\tilde{L}}\right)^2 t\right)\ . \qquad (2.38) \end{aligned}$$

The expression in Eq. (2.37) agrees with the solution of a similar problem in Ref. [66]. For the numerical evaluation only the first 100 terms of the sum have been considered.

We perform corresponding Langevin dynamics simulations according to the technique in Sect. 2.1. In these simulations two particles are restricted to one-dimensional movement by fixing the other components of the position, $y = z = 0$. The passage is defined as the situation when the reactive areas overlap, which is our geometric criterion for an encounter state. Fig. 2.2 shows excellent agreement between the simulation results and the theoretic predictions derived in this



(a) **(b)**

Figure 2.2: (a) Mean first passage time T and (b) selected first passage time distributions f (double-logarithmic) obtained by simulating a system as described in Fig. 2.1 with the following parameters $\eta_0 = R = 1$, $R_c = 1.1$, $\tilde{L} = 2.8$ and time step $\tau = 0.001$. The data points give the simulation results for several relative starting positions $\tilde{x}_0$, while the dotted lines represent the analytic predictions.

section. Due to the chosen simulation method, the first passage time is systematically overestimated, which will be discussed in more detail in the next section.

Because in most situations of interest, the initial condition cannot be controlled, we now proceed by averaging over the initial configurations. Recalling Eq. (2.32), the average mean first passage time $\overline{T}$ is:

$$\overline{T} = \frac{1}{\tilde{L}} \int_0^{\tilde{L}} \frac{1}{D} \left(\tilde{L} - \tilde{x}_0 \right) \tilde{x}_0 = \frac{\tilde{L}^2}{6D} \, . \tag{2.39}$$

Integrating over all possible $\tilde{x}_0$ in Eq. (2.38) we obtain the average first passage time distribution $\overline{f}(t)$:

$$\begin{aligned} \overline{f}(t) &= \sum_{n=0}^{\infty} \frac{2(2n+1)\pi D}{\tilde{L}^2} \left[\frac{1}{\tilde{L}} \int_0^{\tilde{L}} \sin\left(\frac{\pi(2n+1)}{\tilde{L}} \tilde{x}_0 \right) \right] \exp\left(-\frac{D}{2} \left(\frac{\pi(2n+1)}{\tilde{L}} \right)^2 t \right) \\ &= \frac{4D}{\tilde{L}^2} \sum_{n=0}^{\infty} \exp\left(-\frac{D}{2} \left(\frac{\pi(2n+1)}{\tilde{L}} \right)^2 t \right) \, . \end{aligned} \tag{2.40}$$

$\overline{f}(t)$ is an infinite sum of Poisson distributions of different decay times. Therefore, one can expect that for large t, only the slowest decay with $n = 0$ will be important, i.e., $f(t) \sim \exp[-(\pi^2 D/2\tilde{L}^2)t]$ for $t \to \infty$. In Fig. 2.3, not only the exponential decay is compared to the simulation data, but also another feature can be clearly observed: For small t, the distribution shows a power-law behavior $f(t) \sim 1/t^{-1/2}$. It is not trivial to derive this behavior from Eq. (2.40). However, it can be motivated by going over to an integral approximation of the sum. The integral of $\exp[-a(2n+1)^2 t]$ underestimates the sum, which would be represented by a step function in

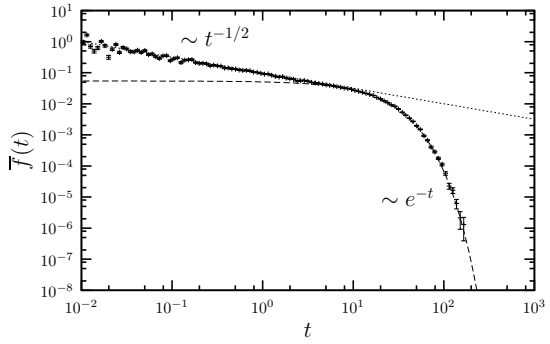


Figure 2.3: Double-logarithmic plot of first passage time distributions of the binding event in the system considered in Fig. 2.2, averaged over the possible initial configurations. The dotted and dashed lines hint the scaling behavior at small and large t respectively.

the integral picture.

$$\sum_{n=0}^{\infty} e^{-a(2n+1)^2 t} > \int_0^{\infty} \mathrm{d}n\, e^{-a(2n+1)^2 t}\,, \quad \text{with} \quad a = \frac{D}{2}\frac{\pi^2}{\tilde{L}^2}\,. \tag{2.41}$$

Since the n dependence of the exponent is not linear, the deviation of the integral representation from this step function is more complicated than a single factor. However, for a reasonably rapid decay, the largest part of the deviation results from the first several "steps" of n, where the discrepancy can still be well described by a factor:

$$\begin{aligned}\sum_{n=0}^{\infty} e^{-a(2n+1)^2 t} &\approx \left(\int_0^{\infty} \mathrm{d}n\, e^{-a(2n+1)^2 t}\right)\left[e^{-at}\left(\int_0^{1} \mathrm{d}n\, e^{-a(2n+1)^2 t}\right)^{-1}\right] \\ &= \frac{\sqrt{\pi}}{4\sqrt{at}}\left(1-\mathrm{erf}(\sqrt{at})\right)\left[\frac{e^{-at}}{\frac{\sqrt{\pi}}{4\sqrt{at}}\left(\mathrm{erf}(3\sqrt{at})-\mathrm{erf}(\sqrt{at})\right)}\right]\,. \end{aligned} \tag{2.42}$$

In the last equation, $\mathrm{erf}(z)$ denotes the error function. Fig. 2.4a illustrates the considerations. The area beneath the solid line would be the exact solution of the sum, the dashed line gives the integral approximation and finally the dotted line is the integral approximation corrected by the factor in squared brackets above. The first order approximation for the error function is $\mathrm{erf}(z) = 2z/\sqrt{\pi}$ [1]. Therefore, a good approximation of Eq. (2.40) for small t is:

$$\overline{f}(t) \approx \frac{4D}{\tilde{L}^2} e^{-at} \frac{1-\mathrm{erf}(\sqrt{at})}{\mathrm{erf}(3\sqrt{at})-\mathrm{erf}(\sqrt{at})} \tag{2.43}$$

$$\approx \frac{4D}{\tilde{L}^2} e^{-at} \frac{\frac{\sqrt{\pi}}{2}-\sqrt{at}}{3\sqrt{at}-\sqrt{at}} \approx \frac{4D}{\tilde{L}^2}\frac{\sqrt{\pi}e^{-at}}{4\sqrt{at}}\,. \tag{2.44}$$

The constant a was defined in Eq. (2.41). For large z, erf$(z) \to 1$. Thus, the limit of the factor with the error functions in Eq. (2.43) can be calculated by differentiating numerator and denominator:

$$\lim_{z\to\infty} \frac{1-\mathrm{erf}(z)}{\mathrm{erf}(3z)-\mathrm{erf}(z)} = \lim_{z\to\infty} \frac{-e^{-z^2}}{3e^{-9z^2}-e^{-z^2}} = 1\,. \tag{2.45}$$

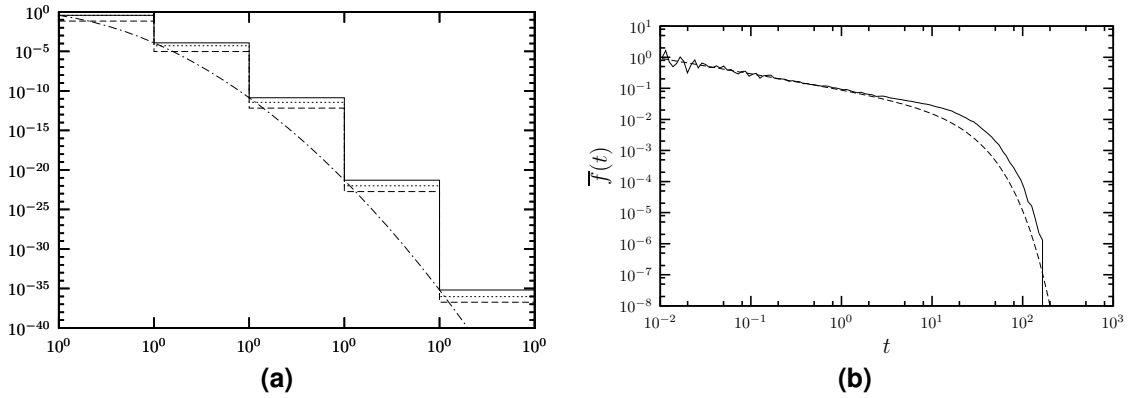


Figure 2.4: (a) Continuous function $\exp[-(2x+1)^2]$ (chain line) is compared to a respective step-function (solid line), whose integration would be equivalent to the sum in Eq. (2.40). The dashed and the dotted lines compare the latter with a step-wise integration of the continuous function, where the dotted line includes the correction-factor for $x \in [0,1]$ as used in Eq. (2.42). (b) Same data as in Fig. 2.3 is compared to the approximate solution obtained in Eq. (2.44).

Therefore, the behavior for large t in Eq. (2.40), which is the slowest decay for $n = 0$, is reproduced by (2.43), as the fraction with the error functions approaches 1 as $t \to \infty$. However, the short-time behavior given in Eq. (2.44) has an additional term $1/\sqrt{t}$, which vanishes for $t \to \infty$. But as the e^{-t} decays much faster, the effect of $1/\sqrt{t}$ can be generally neglected for large t. In this sense, Eq. (2.44) gives a suitable approximation for Eq. (2.40). In Fig. 2.4b the approximation is compared to the simulation data. As expected, there is a deviation for larger t, but the qualitative behavior is well reproduced.

2.2.3 Diffusion between two absorbing spherical shells

We now turn to three dimensions by considering a two-particle system, which is enclosed by an absorbing spherical boundary with radius $\tilde{L} = L + R_c$. If one of the two particles is fixed at the center, the center of mass of the second one $\mathbf{r}$ undergoes diffusion between two absorbing spherical shells of radius $2R_c$ and L (see Fig. 2.5). In this case, the diffusion constant D of the effective process is simply the translational diffusion constant of the mobile particle $D = D_t = 1/(6\pi\eta_0 R)$. The mean first passage time $T(\mathbf{r}_0)$ is the solution of Eq. (2.30) with respect to $\mathbf{r}_0$:

$$\triangle_{\mathbf{r}_0} T(\mathbf{r}_0) = -\frac{2}{D} \,. \tag{2.46}$$

$\mathbf{r}_0$ denotes the initial position of the system at time $t = 0$. Since the particular problem has spherical symmetry, it is $T(\mathbf{r}_0) = T(r_0)$ and Eq. (2.46) can be simplified to an ordinary differential equation by using the Laplacian in spherical coordinates:

$$\frac{1}{r_0^2}\frac{\partial}{\partial r_0} r_0^2 \frac{\partial}{\partial r_0} T(r_0) = -\frac{2}{D} \,, \tag{2.47}$$

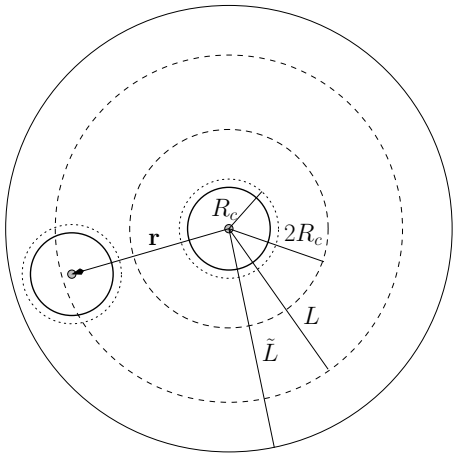


Figure 2.5: A simple model in three dimensions consists of two finite-sized particles, surrounded by an absorbing spherical shell, whose center coincides with one of the two particles, while the other particle can move freely in three dimensions. The two dashed circles denote the effective inner and outer radii, at which the center of the free particle is captured.

whose general solution can be found via twofold integration with respect to r:

$$T(r_0) = -\frac{r_0^2}{3D} + \frac{C_1}{r_0} + C_2 \ . \tag{2.48}$$

The boundary conditions $T(2R_c) = 0$ and $T(L) = 0$ lead to $C_1 = -(4LR_c^2 + 2R_cL^2)$ and $C_2 = L^2 + 4R_c^2 + 2LR_c$, which results in:

$$T(r_0) = \frac{(L - r_0)(r_0 - 2R_c)(L + 2R_c + r_0)}{3Dr_0} \ . \tag{2.49}$$

Following the same steps as in Sect. 2.2.2 one can also attempt to find an expression for the first passage time distribution $f(r_0, t)$. Eq. (2.33) has the same form in three dimensions:

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = \frac{D}{2} \triangle p(\mathbf{r}, t) \ , \tag{2.50}$$

and can again be rewritten using the spherical symmetry:

$$\frac{\partial p(r, t)}{\partial t} = \frac{D}{2} \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} p(r, t) \ . \tag{2.51}$$

According to [113] a possible solution for the heat equation with spherical symmetry is:

$$p(r, t) = A \frac{1}{r} \exp\left(-\mu^2 Dt/2\right) \cos\left(\mu r + B\right) + C \ . \tag{2.52}$$

Using again the boundary conditions already stated above gives the following full solution:

$$p(r, t) = \sum_{n=1}^{\infty} A_n \frac{1}{r} \exp\left(-\frac{D}{2}\left(\frac{\pi n}{L - 2R_c}\right)^2 t\right) \sin\left(\left(\frac{\pi n}{L - 2R_c}\right)(r - 2R_c)\right) \ . \tag{2.53}$$

Because of the factor $1/r$, this solution is not periodic anymore and cannot be treated like a Fourier series. Comparing Eq. (2.53) to the solution given in Ref.

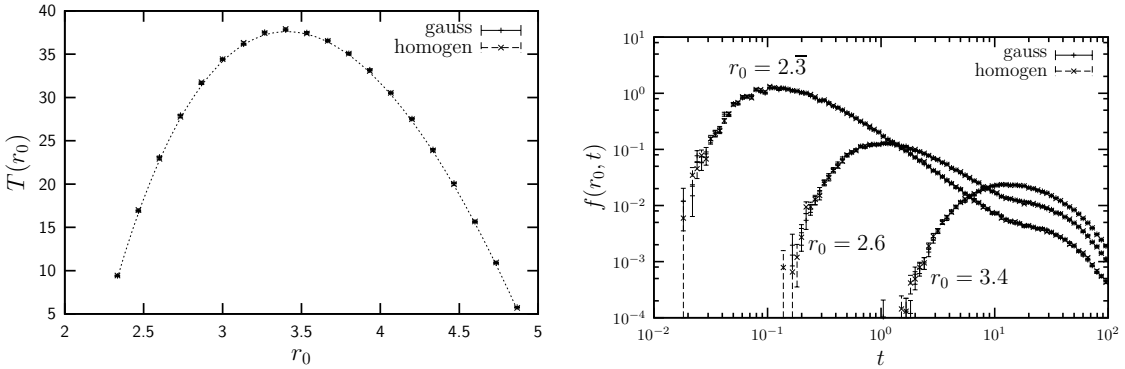


Figure 2.6: (a) Mean first passage time T and (b) selected first passage time distributions f (double-logarithmic) obtained by simulating a system as described in Fig. 2.5 with the following parameters $\eta_0 = R = 1$, $R_c = 1.1$, $L = 5.0$ and time step $\Delta t = 0.001$. The data points give the simulation results for several relative starting positions r_0, while the dotted line in (a) represents the analytic prediction for T. In both pictures, data obtained by simulating with random numbers out of a homogeneous (represented by the bar symbols) and a Gaussian distribution (cross symbols) are compared.

[34], one still finds a similar expression for the evaluation of A_n, assuming that the initial condition is again $p(\tilde{r}, 0|\tilde{r}_0, 0) = \delta(\tilde{r} - \tilde{r}_0)$:

$$\begin{aligned} A_n &= \frac{2}{L - 2R_c} \int_{2R_c}^{L} \mathrm{d}r'\, p(\tilde{r}, 0|\tilde{r}_0, 0) r' \sin\left(\left(\frac{\pi n}{L - 2R_c}\right)(r' - 2R_c)\right) \\ &= \frac{2r_0}{L - 2R_c} \sin\left(\left(\frac{\pi n}{L - 2R_c}\right)(r_0 - 2R_c)\right) \ . \end{aligned} \tag{2.54}$$

It is not straight-forward to check whether the A_n are normalized. The first passage time distribution $f(r_0, t) = -\int \mathrm{d}r'\, \dot{p}(r', t)$ contains an infinite sum over an integral which can only be evaluated numerically:

$$\begin{aligned} f(r_0, t) = \sum_{n=0}^{\infty} \mu_n^2 \frac{D r_0}{(L - 2R_c)} \sin\left(\mu_n (r_0 - 2R_c)\right) \times \\ \exp\left(-\frac{D}{2}\mu_n^2 t\right) \left[\int_{2R_c}^{L} \mathrm{d}r' \frac{1}{r'} \sin\left(\mu_n (r' - 2R_c)\right)\right] \ , \end{aligned} \tag{2.55}$$

with $\mu_n = \pi n/(L - 2R_c)$. The integral over $\sin(r')/r'$ has a bad convergence due to the alternating behavior. Thus, the expression in Eq. (2.55) is not really applicable for numerical calculations even in very sophisticated numerical calculation systems.

Again, we perform corresponding computer simulations to check the results against the prediction in Eq. (2.49) on the one hand, and to verify two technical details of the simulation. First, the simulation of a Langevin dynamics uses

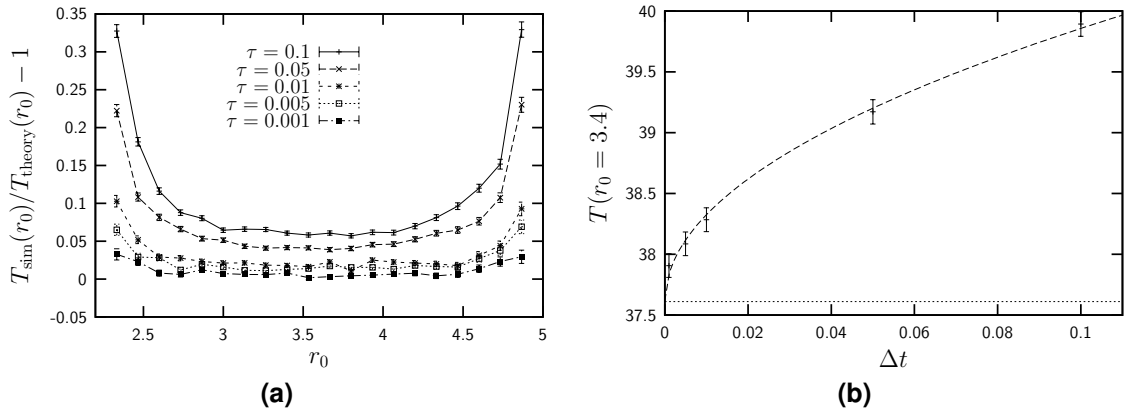


Figure 2.7: (a) Relative deviation of the mean first passage time T measured in the simulations as described in Fig. 2.6 from the analytical prediction for different time steps Δt. (b) A square-root fit (dashed line) to the measured mean first passage time T for different time steps Δt demonstrates the scaling of the systematic error. The dotted line denotes the analytic result.

a technical representation of thermal noise. But as stated in Sect. 2.1, for the particularly chosen simulation scheme the usage of Gaussian and equally distributed random numbers should be equivalent. Fig. 2.6 compares the simulation results of the considered three-dimensional, spherically bounded system for both cases. The data from both methods agree except for statistical deviations. The prediction of the mean first passage time is also well reproduced.

As already mentioned in the last section, the first passage times t are systematically overestimated with the chosen algorithm. Particularly, for a fixed time step Δt in the simulation scheme an error of order $\sim \sqrt{\Delta t}$ arises. Fig. 2.7a shows the relative error of $T = \bar{t}$ with respect to the analytic prediction for different time steps Δt and the homogeneous random number implementation. Indeed, the error is always positive and increases with Δt. At $r_0 \to 2R_c$ and $r_0 \to \tilde{L}$, T decreases and thus the relative error gets larger. In Fig. 2.7b the values of T are plotted with respect to the time step Δt. This demonstrates the square-root behavior of $T_{\mathrm{sim}}(\Delta t)$. The fit agrees very well with the theoretic prediction at $\Delta t \to 0$.

2.3 Evaluation of hydrodynamic properties

Although we do not incorporate hydrodynamic interactions in our model in this work, hydrodynamic properties of the single particles or also partially formed protein complexes are considered via the full 6×6 anisotropic diffusion matrix. In the case of a single protein we will ask for the mobility matrix of one object whose atomistic structure is known, e.g., from x-ray scattering experiments or NMR studies. In dynamic systems with several proteins where bonds are formed and removed all the time, usually no structures are available yet. For both situations a recipe is needed to estimate the mobility matrix.

A promising approach that has been published recently [16] involves several

sophisticated techniques to determine the surface of a protein, make a corresponding triangulation and integrate the hydrodynamic impact of the solvent over the triangles. However, a less involved and widely used approach has been developed mainly by de la Torre and coworkers [25, 32, 33, 39, 41, 76]. With this method it is possible to calculate the mobility matrix of a rigidly bound aggregate of N spheres with respect to both translational and rotational diffusion. Therefore, it intrinsically enables to evaluate the hydrodynamic properties of clusters of generic model proteins. Furthermore, one can think of several approaches to fill up the space occupied by a protein structure with small beads and then use the so obtained aggregate as input for the method. Different models are discussed and compared in Refs. [32, 40]. The basic approach is to put spheres of a certain size at the position of any atom except hydrogens. The volume of these spheres effectively models a fixed hydration shell. This construct is then filled up with smaller, densely packed but non-overlapping beads. Since the hydrodynamic properties of a rigid body are determined by its outer boundary only, a shell of these small spheres can be generated by deleting all spheres which have a maximum number of possible neighbors. The calculation of the diffusion matrix with the method explained in this section is carried out for systematically decreased sizes of the small beads. Thus, the detail of modeling is stepwise increased and the result should in all practical cases converge against reasonable values. The authors provide a freely available software called `HYDROPRO` [40], which uses the structure of a protein obtained from the protein data bank (PDB), e.g., to directly calculate the diffusion matrix. Several system properties are implicitly contained in the diffusion matrix. Throughout this work – unless otherwise stated – we use the ambient temperature $T_a = 293\mathrm{K}$ and for the density and dynamic viscosity of the solvent we choose the respective parameters of water $\rho = 1\mathrm{g/cm^3}$ and $\eta_0 = 10^{-3}\mathrm{Pa\,s}$.

The fundamental concept of de la Torre's method is to consider the free motion of each of the spheres. It is affected by the presence of the other spheres due to hydrodynamic interaction. Finally, the constraints arising from the rigid structure of the aggregate are applied. In this way, general rules can be derived of how to include the pairwise hydrodynamic interactions into a global mobility matrix. The details will be explained in the following.

2.3.1 Hydrodynamic interaction in a system of many spheres

The treatment of de la Torre is based on the work of Brenner and O'Neill [29]. However, we basically follow Ref. [33] in our description and notation. Consider the system of N spheres with linear velocity $\mathbf{u}_i^P$ referred to some point P_i and angular velocities ω_i, $i \in [1, N]$. Each sphere experiences a frictional force $\mathbf{F}_i$ and a frictional torque $\mathbf{T}_i^P$ when moving, arising from the presence of the other spheres. The frictional impact of one sphere i on another one j can be expressed in 3x3 matrices ζ_{ij}, accounting for the translational (tt), rotational (rr) and translational-rotational (tr/rt) coupling, which are not specified any further at the moment. In

the overdamped limit, this leads to the following equations:

$$\mathbf{F}_i = \sum_{j=1}^{N} \zeta_{ij}^{tt} \mathbf{u}_j^P + \sum_{j=1}^{N} \zeta_{ij}^{tr} \boldsymbol{\omega}_j \ , \tag{2.56}$$

$$\mathbf{T}_i^P = \sum_{j=1}^{N} \zeta_{ij}^{rt} \mathbf{u}_j^P + \sum_{j=1}^{N} \zeta_{ij}^{rr} \boldsymbol{\omega}_j \ . \tag{2.57}$$

One can also generalize this to equations with supervectors including all velocities ($\boldsymbol{\mathcal{U}}^P$, $\boldsymbol{\mathcal{O}}$), forces ($\boldsymbol{\mathcal{F}}$) or torques ($\boldsymbol{\mathcal{T}}^P$) of the system:

$$\begin{pmatrix} \boldsymbol{\mathcal{F}} \\ \boldsymbol{\mathcal{T}}^P \end{pmatrix} = \begin{pmatrix} \zeta^{tt} & \zeta^{tr} \\ \zeta^{rt} & \zeta^{rr} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{U}}^P \\ \boldsymbol{\mathcal{O}} \end{pmatrix} \ . \tag{2.58}$$

The grand friction matrix should be symmetric, therefore some relationships can be presumed: $\zeta_{ij}^{tt} = \zeta_{ji}^{tt}$, $\zeta_{ij}^{rr} = \zeta_{ji}^{rr}$ and $\zeta_{ij}^{tr} = \zeta_{ji}^{rt}$. The relationship in Eq. (2.58) can also be inverted to calculate the velocities out of the forces and torques with respect to a grand mobility matrix μ:

$$\begin{pmatrix} \boldsymbol{\mathcal{U}} \\ \boldsymbol{\mathcal{O}}^P \end{pmatrix} = \begin{pmatrix} \mu^{tt} & \mu^{tr} \\ \mu^{rt} & \mu^{rr} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{F}}^P \\ \boldsymbol{\mathcal{T}} \end{pmatrix} \ , \quad \text{with} \quad \begin{pmatrix} \mu^{tt} & \mu^{tr} \\ \mu^{rt} & \mu^{rr} \end{pmatrix} = \begin{pmatrix} \zeta^{tt} & \zeta^{tr} \\ \zeta^{rt} & \zeta^{rr} \end{pmatrix}^{-1} \ . \tag{2.59}$$

In Eqs. (2.56) and (2.57) the constraints can now be taken into account, which come from the assumption that the sum of all spheres forms a rigid body and thus no relative motion of the spheres is allowed. The whole body shall have a translational velocity $\mathbf{u}^O$ referred to some origin O and an angular velocity $\boldsymbol{\omega}$. Setting $P_i = O$, the velocities of the single spheres can be evaluated to $\mathbf{u}_i^P = \mathbf{u}^O + \boldsymbol{\omega} \times \mathbf{r}_i$ and $\boldsymbol{\omega}_i = \boldsymbol{\omega}$, where $\mathbf{r}_i$ denotes the distance vector of the center of sphere i to the origin O. Now it is possible to calculate the frictional force and torque of the whole rigid body out of $\mathbf{F}_i$ and $\mathbf{T}_i^P$:

$$\mathbf{F} = \Xi^{tt} \mathbf{u}^O + \Xi_O^{tr} \omega = \sum_{i=1}^{N} \mathbf{F}_i \ , \tag{2.60}$$

$$\mathbf{T}^O = \Xi_O^{rt} \mathbf{u}^O + \Xi_O^{rr} \omega = \sum_{i=1}^{N} \left(\mathbf{F}_i^P + \mathbf{r}_i \times \mathbf{F}_i \right) \ . \tag{2.61}$$

Generally, cross products $\mathbf{r}_i \times \boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}} \times \mathbf{r}_i$ can also be expressed as a matrix product $\mathbb{A}_i \boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}} \mathbb{A}_i$ respectively, with

$$\mathbb{A}_i = \begin{pmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{pmatrix} \ . \tag{2.62}$$

By inserting Eqs. (2.56) and (2.57) into Eqs. (2.60) and (2.61) and using Eq. (2.62), it is shown in Ref. [33] that:

$$\Xi^{tt} = \sum_i \sum_j \zeta_{ij}^{tt} \,, \tag{2.63}$$

$$\Xi_O^{tr} = \sum_i \sum_j \left(-\zeta_{ij}^{tt} \mathbb{A}_j + \zeta_{ij}^{tr} \right) \,, \tag{2.64}$$

$$\Xi_O^{rt} = \sum_i \sum_j \left(\zeta_{ij}^{rt} + \mathbb{A}_i \zeta_{ij}^{tt} \right) \,, \tag{2.65}$$

$$\Xi_O^{rr} = \sum_i \sum_j \left(\zeta_{ij}^{rr} - \zeta_{ij}^{rt} \mathbb{A}_j + \mathbb{A}_i \zeta_{ij}^{tr} - \mathbb{A}_i \zeta_{ij}^{tt} \mathbb{A}_j \right) \,. \tag{2.66}$$

Now the question is how to obtain the friction tensors of the pairwise hydrodynamic interaction ζ_{ij}. Usually, the friction tensors cannot be directly obtained, but several methods have been developed to estimate the diffusive interaction μ_{ij}. In Ref. [32] the Kirkwood-Riseman-like [84] treatments are utilized, which are discussed in-depth in Ref. [33] (section C) amongst others. Generally, these type of methods neglect the rotational and translation-rotational coupling, i.e. $\mu_{ij}^{rr} \equiv \mu_{ij}^{tr} \equiv \mu_{ij}^{rt} \equiv \mathbb{O}$. However, in Ref. [41] a convenient way to include a volume correction is described, which is equivalent to a zeroth order contribution to the rotational coupling and will be discussed later. Basically, the translational "self-coupling" of a bead is accounted for by the reciproke of the standard translational friction coefficient $\mu_{ii}^{tt} = (6\pi\eta_0\sigma_i)^{-1}$, where η_0 is the viscosity of the medium and σ_i is the radius of bead i. The hydrodynamic interaction of two spherical beads with distance vector $\mathbf{R}_{ij}$ is described by means of the Oseen tensor [83]:

$$\mu_{ij}^{tt} = \begin{cases} \left(8\pi\eta_0 R_{ij}^3\right)^{-1} \left(R_{ij}^2 \left(\mathbb{I} + \mathbb{P}_{ij}\right)\right) & \text{if } i \neq j \\ \left(6\pi\eta_0\sigma_i\right)^{-1} \mathbb{I} & \text{if } i = j \end{cases} \,, \quad \text{with } \mathbb{P}_{ij} = \mathbf{R}_{ij} \otimes \mathbf{R}_{ij} / R_{ij}^2 \,, \tag{2.67}$$

where $\mathbf{R}_{ij} \otimes \mathbf{R}_{ij}$ denotes a dyadic product, $\mathbb{I}$ is the unit matrix and $R_{ij} = |\mathbf{R}_{ij}|$. A common variant of a modified Oseen tensor for this type of problem was introduced by Rotne, Prager [121] and Yamakawa [163] for two spheres of equal radius. De la Torre and Bloomfield used a more general expression for spheres of different radii σ_i and σ_j [39]:

$$\mu_{ij}^{tt} = \begin{cases} \left(8\pi\eta_0 R_{ij}^3\right)^{-1} \left(R_{ij}^2 \left(\mathbb{I} + \mathbb{P}_{ij}\right) + \left(\sigma_i^2 + \sigma_j^2\right) \left(\frac{1}{3}\mathbb{I} - \mathbb{P}_{ij} \right) \right) & \text{if } i \neq j \\ \left(6\pi\eta_0\sigma_i\right)^{-1} \mathbb{I} & \text{if } i = j \end{cases} \,. \tag{2.68}$$

Note that for overlapping spheres ($\sigma_i + \sigma_j > R_{ij}$) at least the non-diagonal entries in Eq. (2.68) can get negative which possibly conflicts the requirement of the diffusion matrix μ_{ij}^{tt} to be positively definite. Rotne and Prager discuss this and give an alternative expression to prevent this problem. However, Eq. (2.68) is only valid for non-overlapping spheres. Considering only μ^{tt} to be a non-zero matrix

$\mu^{tt} \neq \mathbb{O}$ it is also $\zeta^{rr} = \zeta^{tr} = \zeta^{rt} = \mathbb{O}$ and thus Eqs. (2.63)–(2.66) simplify to:

$$\Xi^{tt} = \sum_i \sum_j \zeta_{ij}^{tt} \,, \tag{2.69}$$

$$\Xi_O^{tr} = -\sum_i \sum_j \zeta_{ij}^{tt} \mathbb{A}_i \,, \tag{2.70}$$

$$\Xi_O^{rt} = \sum_i \sum_j \mathbb{A}_i \zeta_{ij}^{tt} \,, \tag{2.71}$$

$$\Xi_O^{rr} = -\sum_i \sum_j \mathbb{A}_i \zeta_{ij}^{tt} \mathbb{A}_j \,. \tag{2.72}$$

Note that in Ref. [32] there is a spelling mistake in the definition of Ξ_O^{rr} – Eq. (25) (which is equivalent to our Eq. (2.72)) lacks a minus sign.

Center of diffusion

Although the grand mobility matrix μ does not depend on the chosen coordinate system in which it is calculated as only relative positions go into Eq. (2.68), it is clear from Eqs. (2.69)–(2.72) that Ξ_O is coordinate dependent because of the products with $\mathbb{A}$. In Ref. [76] the authors explain how to choose the correct point of reference for a certain system. The basic idea is that finally either the friction matrix or the diffusion matrix, depending on which quantity is wanted, should be symmetric. Particularly, they point out that generally the certain point $\mathbf{O} = \mathbf{R}$ which leads Ξ_O to be symmetric (the "center of friction") is different from the "center of diffusion" $\mathbf{O} = \mathbf{D}$, where $\mathbb{D}_O$ is symmetric. They further state that $\operatorname{Tr} \Xi_O^{rr}$ as well as $\operatorname{Tr} \mathbb{D}_O^{tt}$ have their minimum at the respective center.

Another interesting fact is that although the *translational* friction matrix Ξ^{tt} is independent of the chosen center, it is the *rotational* diffusion matrix $\mathbb{D}^{rr}$, for which this independence is valid. This point is rather counterintuitive at first glance, since $\mathbb{D}^{tt}$ is of course related to Ξ^{tt}. However, according to Ref. [76] the origin independence of Ξ^{tt} reflects the fact that the overall force $\mathbf{F}$ should also be independent of the origin, contrary to the overall torque $\mathbf{T}^O$. Instead, $\mathbb{D}^{tt}$ is connected to the linear velocity $\mathbf{u}^O$, which depends on the origin, while $\mathbb{D}^{rr}$ is connected to the angular velocity ω, which does not. This can be accounted for a phenomenological argument.

In Ref. [25] the calculation of the center of friction is motivated, presuming Ξ_O has been evaluated before with respect to some arbitrary origin $\mathbf{O}$. In Ref. [32] the analogous expression for the center of diffusion $\mathbf{D}$ can be found in dependence of $\mathbb{D}_O$:

$$\mathbf{r}_{DO} = \mathbf{D} - \mathbf{O} = \begin{pmatrix} \mathbb{D}_{yy}^{rr} + \mathbb{D}_{zz}^{rr} & -\mathbb{D}_{xy}^{rr} & -\mathbb{D}_{xz}^{rr} \\ -\mathbb{D}_{xy}^{rr} & \mathbb{D}_{xx}^{rr} + \mathbb{D}_{zz}^{rr} & -\mathbb{D}_{yz}^{rr} \\ -\mathbb{D}_{xz}^{rr} & -\mathbb{D}_{yz}^{rr} & \mathbb{D}_{xx}^{rr} + \mathbb{D}_{yy}^{rr} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{D}_{yz}^{rt} - \mathbb{D}_{zy}^{rt} \\ \mathbb{D}_{zx}^{rt} - \mathbb{D}_{xz}^{rt} \\ \mathbb{D}_{xy}^{rt} - \mathbb{D}_{yx}^{rt} \end{pmatrix} . \tag{2.73}$$

It is possible to shift a diffusion or friction matrix calculated with respect to some point $\mathbf{O}$ to another center. The required identities are given in Ref. [25].

Shifting the diffusion matrix to the center of diffusion $\mathbf{D}$ from Eq. (2.73) is done by the following rules:

$$\mathbb{D}_D^{tt} = \mathbb{D}_O^{tt} - \mathbb{A}_{DO}\mathbb{D}_O^{rr}\mathbb{A}_{DO} + \mathbb{D}_{\mathbf{O}}^{rt\dagger}\mathbb{A}_{DO} - \mathbb{A}_{DO}\mathbb{D}_O^{rt} \,, \tag{2.74}$$

$$\mathbb{D}_D^{rt} = \mathbb{D}_O^{rt} + \mathbb{D}_O^{rr}\mathbb{A}_{DO} \,, \tag{2.75}$$

$$\mathbb{D}_D^{rr} = \mathbb{D}_O^{rr} \,. \tag{2.76}$$

Here, $\mathbb{A}_{DO}$ corresponds to the definition of $\mathbb{A}_i$ in Eq. (2.62) with $\mathbf{r}_{DO}$ from Eq. (2.73) as the argument.

Volume correction

In Sect. 2.3.1 the estimation of μ_{ij} was described, mentioning that usually all components except μ^{tt} are neglected. Considering a single sphere and recalling Eqs. (2.70)–(2.72) this would mean that all components of the friction tensor would vanish except the translational part $\boldsymbol{\Xi}^{tt} \neq \mathbf{0}$. This is not only an obvious mistake, but it also causes a computational problem, because a matrix of the form

$$\boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{\Xi}^{tt} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \tag{2.77}$$

is singular ($\det \boldsymbol{\Xi} = 0$) and thus cannot be inverted. Therefore, also the diffusion tensor would be undefined. Apart from that, the estimation of the intrinsic viscosity is also affected by this failure. This problem was tackled in Ref. [41]. It was suggested to introduce a volume correction for the rotational properties by adding a rotational friction coefficient to the trace of $\boldsymbol{\Xi}^{rr}$, which is equal to that of a virtual sphere of the volume of all spheres in the system:

$$\boldsymbol{\Xi}_{\mathrm{corr}}^{rr} = \boldsymbol{\Xi}^{rr} + 6\eta_0 V \mathbb{I} \,, \qquad V = \frac{4}{3}\pi \sum_{i=1}^{N} \sigma_i^3 \,. \tag{2.78}$$

As it is shown in Ref. [32], this volume correction does not necessarily improve the expected results, especially in the case when structures are modeled by a number of smaller spheres. Although, since it removes the computational problems arising by the appearance of singular friction matrices, it is still valuable to include it into the calculations.

Overall procedure

Summarizing all considerations the method consists of the following steps:

1. evaluate Eq. (2.68) to calculate the grand translational mobility matrix μ^{tt},
2. invert μ^{tt} to get $\zeta^{tt} = (\mu^{tt})^{-1}$,
3. calculate $\boldsymbol{\Xi}_O$ out of ζ^{tt} utilizing Eqs. (2.69)–(2.72) with respect to some arbitrary point $\mathbf{O}$ (e.g. the origin $\mathbf{0}$),

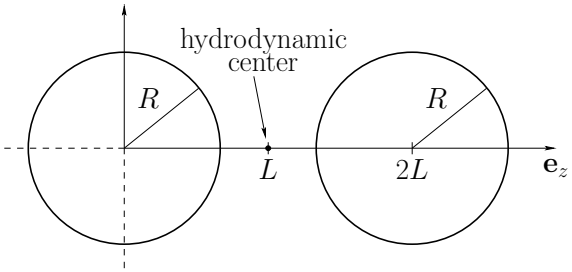


Figure 2.8: Scheme of the considered exemplary geometry. You see two spheres arranged along the $\mathbf{e}_z$ axis, one centered at the origin, the other translated by $2L\mathbf{e}_z$. As it will be shown in the calculation, the hydrodynamic center is located at $L\mathbf{e}_z$.

4. include the volume correction from Eq. (2.78),
5. invert $\boldsymbol{\Xi}$ to get $\mathbb{D}_O = k_B T_a \boldsymbol{\Xi}^{-1}$,
6. use Eq. (2.73) to find the center of diffusion $\mathbf{D}$,
7. shift $\mathbb{D}_O$ to the center of diffusion with the prescript given in Eqs. (2.74)–(2.76) to obtain the final result: the estimate of the full diffusion matrix of a system of N arbitrarily sized spheres $\mathbb{D}_D$.

2.3.2 Simple example – the dumbbell

To demonstrate the procedure and to get an idea of its consequences, we now calculate the diffusion coefficient of a dumbbell consisting of two equal-sized spheres of radius R fixed to each other at a distance $2L$ via the previously described method. Fig. 2.8 illustrates the considered geometry. We assume that the locations of the spheres are $\mathbf{r}_1 = \mathbf{0}$ and $\mathbf{r}_2 = 2L\mathbf{e}_z$, which makes $\mathbf{R}_{12} = -2L\mathbf{e}_z$. Therefore, according to Eq. (2.68) μ^{tt} has the form:

$$\mu^{tt} = \begin{pmatrix} a & 0 & 0 & b & 0 & 0 \\ 0 & a & 0 & 0 & b & 0 \\ 0 & 0 & a & 0 & 0 & c \\ b & 0 & 0 & a & 0 & 0 \\ 0 & b & 0 & 0 & a & 0 \\ 0 & 0 & c & 0 & 0 & a \end{pmatrix} , \tag{2.79}$$

$$\begin{aligned} \text{with} \quad a &= \frac{1}{6\pi\eta_0 R} , \quad b = \frac{1}{8\pi\eta_0 (2L)^3} \left((2L)^2 + \frac{2}{3}R^2 \right) , \\ c &= \frac{1}{8\pi\eta_0 (2L)^3} \left(2(2L)^2 - \frac{4}{3}R^2 \right) . \end{aligned} \tag{2.80}$$

Inverting a matrix of the type in Eq. (2.79) generally results in:

$$\left(\mu^{tt}\right)^{-1} = \begin{pmatrix} a' & 0 & 0 & c' & 0 & 0 \\ 0 & a' & 0 & 0 & c' & 0 \\ 0 & 0 & b' & 0 & 0 & d' \\ c' & 0 & 0 & a' & 0 & 0 \\ 0 & c' & 0 & 0 & a' & 0 \\ 0 & 0 & d' & 0 & 0 & b' \end{pmatrix} , \tag{2.81}$$

$$\text{with} \quad a' = \frac{a}{a^2 - b^2} , \quad b' = \frac{a}{a^2 - c^2} , \quad c' = \frac{b}{b^2 - a^2} , \quad d' = \frac{c}{c^2 - a^2} . \tag{2.82}$$

The next step is the evaluation of $\boldsymbol{\Xi}$ according to Eqs. (2.69)–(2.72). Because sphere number 1 is located at the origin it holds $\mathbb{A}_1 = 0$. The volume correction (2.78) is $\boldsymbol{\Xi}^{rr}_{\text{corr}} = \boldsymbol{\Xi}^{rr} + v_c \mathbb{I}$ for this system, with $v_c = 16\pi\eta_0 R^3$. Summing up all contributions, $\boldsymbol{\Xi}$ has the following form:

$$\boldsymbol{\Xi} = \begin{pmatrix} 2(a'+c') & 0 & 0 & 0 & 2L(a'+c') & 0 \\ 0 & 2(a'+c') & 0 & -2L(a'+c') & 0 & 0 \\ 0 & 0 & 2(b'+d') & 0 & 0 & 0 \\ 0 & -2L(a'+c') & 0 & 4a'L^2 + v_c & 0 & 0 \\ 2L(a'+c') & 0 & 0 & 0 & 4a'L^2 + v_c & 0 \\ 0 & 0 & 0 & 0 & 0 & v_c \end{pmatrix} . \tag{2.83}$$

We calculate the diffusion matrix with respect to the initially assumed point of reference by inverting $\boldsymbol{\Xi}$:

$$\mathbb{D}_O = k_B T_a \begin{pmatrix} a^* b^* & 0 & 0 & 0 & -b^* L & 0 \\ 0 & a^* b^* & 0 & b^* L & 0 & 0 \\ 0 & 0 & \dfrac{1}{2(b'+d')} & 0 & 0 & 0 \\ 0 & b^* L & 0 & b^* & 0 & 0 \\ -b^* L & 0 & 0 & 0 & b^* & 0 \\ 0 & 0 & 0 & 0 & 0 & \dfrac{1}{v_c} \end{pmatrix} , \tag{2.84}$$

$$\text{with} \quad a^* = \frac{4a'L^2 + v_c}{2(a'+c')} , \quad b^* = \frac{1}{2(a'-c')L^2 + v_c} . \tag{2.85}$$

Now Eq. (2.73) can be evaluated, to obtain the center of diffusion:

$$\mathbf{r}_{DO} = \begin{pmatrix} \frac{1}{v_c} + b^* & 0 & 0 \\ 0 & \frac{1}{v_c} + b^* & 0 \\ 0 & 0 & 2b^* \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ 2b^* L \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ L \end{pmatrix} . \tag{2.86}$$

Because of the simplicity of the example, the center of diffusion coincides with the center of mass. As already mentioned, for more complex examples hydrodynamic screening will occur, which diminishes the contributions of spheres lying inside the cluster, and thus $\mathbf{r}_{DO}$ is different from the center of mass in general. The knowledge of $\mathbf{r}_{DO}$ finally leads to the diffusion matrix with respect to the center of diffusion by shifting $\mathbb{D}_O$ according to Eqs. (2.74)–(2.76):

$$\mathcal{D}_D = \begin{pmatrix} \mathcal{D}^t_\perp & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathcal{D}^t_\perp & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{D}^t_\parallel & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathcal{D}^r_\perp & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathcal{D}^r_\perp & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathcal{D}^r_\parallel \end{pmatrix} , \tag{2.87}$$

with

$$\mathcal{D}^t_\parallel = k_B T_a \frac{1}{2(b' + d')} = D^t_{\mathrm{sph}} \left(\frac{1}{2} + \frac{3}{8} l^{-1} - \frac{1}{16} l^{-3} \right) , \tag{2.88}$$

$$\mathcal{D}^t_\perp = k_B T_a b^* (a^* - L^2) = D^t_{\mathrm{sph}} \left(\frac{1}{2} + \frac{3}{16} l^{-1} + \frac{1}{32} l^{-3} \right) , \tag{2.89}$$

$$\mathcal{D}^r_\parallel = k_B T_a \frac{1}{v_c} = D^r_{\mathrm{sph}} \left(\frac{1}{2} \right) , \tag{2.90}$$

$$\mathcal{D}^r_\perp = k_B T_a b^* = D^r_{\mathrm{sph}} \left(\frac{1}{2} + \frac{6}{-12 - 16 l^{-2} + 6 l^{-3} + l^{-5}} \right) . \tag{2.91}$$

Here, $\mathbb{D}_D$ is expressed in terms of the scaled distance of the spheres from the center of mass $l = L/R$. The indices $\mathcal{D}_\parallel$ and $\mathcal{D}_\perp$ denote the coefficients with respect to the direction parallel and perpendicular to the symmetry axis of the dumbbell $\mathbf{e}_z$. To make it easier to understand the result, it is expressed in terms of the known diffusion coefficients for a single sphere of radius R [48]:

$$D^t_{\mathrm{sph}} = \frac{k_B T_a}{6\pi\eta_0 R} , \qquad D^r_{\mathrm{sph}} = \frac{k_B T_a}{8\pi\eta_0 R^3} . \tag{2.92}$$

The two interesting limits are $l = 1$, when the spheres touch each other, and $l \to \infty$, when the spheres are far apart from each other. First it is obvious that $\mathcal{D}^r_\parallel$ does not exhibit an l dependence and thus the dumbbell is assumed to be always half as mobile as the single sphere with respect to rotations around $\mathbf{e}_z$ in this method. Regarding translational diffusion for large l all terms $l^{-\alpha}$ vanish and the

friction of both beads is fully accounted for, which means that the whole dumbbell is half as mobile as one single sphere $\mathcal{D}^t_\parallel = \mathcal{D}^t_\perp = D^t_{\mathrm{sph}}/2$. If l gets smaller, the beads experience a pairwise hydrodynamic screening which decreases the friction, i.e., increases the mobility up to some value which is still below the single sphere. Particularly, it is reasonable that translation in the $\mathbf{e}_z$ direction is favored compared to perpendicular translation, where the spheres are more exposed to fluid flow. The rotational diffusion perpendicular to $\mathbf{e}_z$ is always more than a factor 2 smaller compared to a single sphere and even vanishes for large separations $l \to \infty$. Considering the definition of torque $\mathbf{T} = \mathbf{r} \times \mathbf{F}$ helps to explain this behavior physically. Let ω be the velocity of rotation of the whole dumbbell around $\mathbf{e}_y$. This actually means that each of the two beads move with a speed $u_{1/2} = \pm\omega L$ and hence experiences the force $F = \pm(\zeta^t(L) - \zeta^t_c(L))\omega L$. Here, ζ^t denotes a coefficient determining the screened friction of the bead itself and ζ^t_c denotes the coupling to the oppositely moving bead. For increasing L it is a reasonable assumption that $\zeta^t(L)$ and $\zeta^t_c(L)$ converge to some finite value, which gives $F \sim \omega L$. In particular, the coupling as well as the screening will vanish for large separations, i.e. $\zeta^t_c(L) \to 0$ and $\zeta^t(L) \to k_B T_a / D^t_{\mathrm{sph}}$ with $L \to \infty$. The torque on the dumbbell arising from these forces is $T = L \times F \sim \omega L^2$, which clearly diverges. The rotational friction coefficient is $\zeta^r = T/\omega$, it is apparent that the rotational mobility vanishes $\mu_r \sim 1/\zeta^r \sim L^{-2}$.

Comparison with analytic results

The hydrodynamic friction of a system of two spheres has been studied analytically [38, 67, 143, 156], which enables us to compare the previously derived approximate results to exact values. In all cases, the system is described in special curvilinear coordinates ξ, η as introduced by Stimson and Jeffery [143], which are connected to cylindrical coordinates z, r:

$$\xi + \imath\eta = \ln \frac{r + \imath(z+a)}{r + \imath(z-a)} \,, \tag{2.93}$$

$$r = a\frac{\sin\eta}{\cosh\xi - \cos\eta} \,, \quad z = a\frac{\sinh\xi}{\cosh\xi - \cos\eta} \,. \tag{2.94}$$

The surface of a sphere on the z-axis is then an iso surface with $\xi = \xi_0$. Davis gives a general recipe to calculate the necessary constants α, β and a for a system of two spheres with radii R_1, R_2 and center-to-center distance $2L$:

$$a = \sqrt{x_1^2 - R_1^2} = \sqrt{x_1^2 - R_2^2} \,, \tag{2.95}$$

$$\alpha = \ln(z_1 + a) - \ln R_1 \,, \tag{2.96}$$

$$\beta = \ln(z_2 + a) - \ln R_2 \,, \tag{2.97}$$

where $z_1 = (4L^2 + R_1^2 - R_2^2)/(4L)$ and $-z_2 = -(4L^2 - R_1^2 + R_2^2)/(4L)$ denote the center positions. The two spheres are then represented by $\xi = \alpha$ and $\xi = -\beta$. From the above equations we see that in case of equal sized beads $R_1 = R_2 = R$

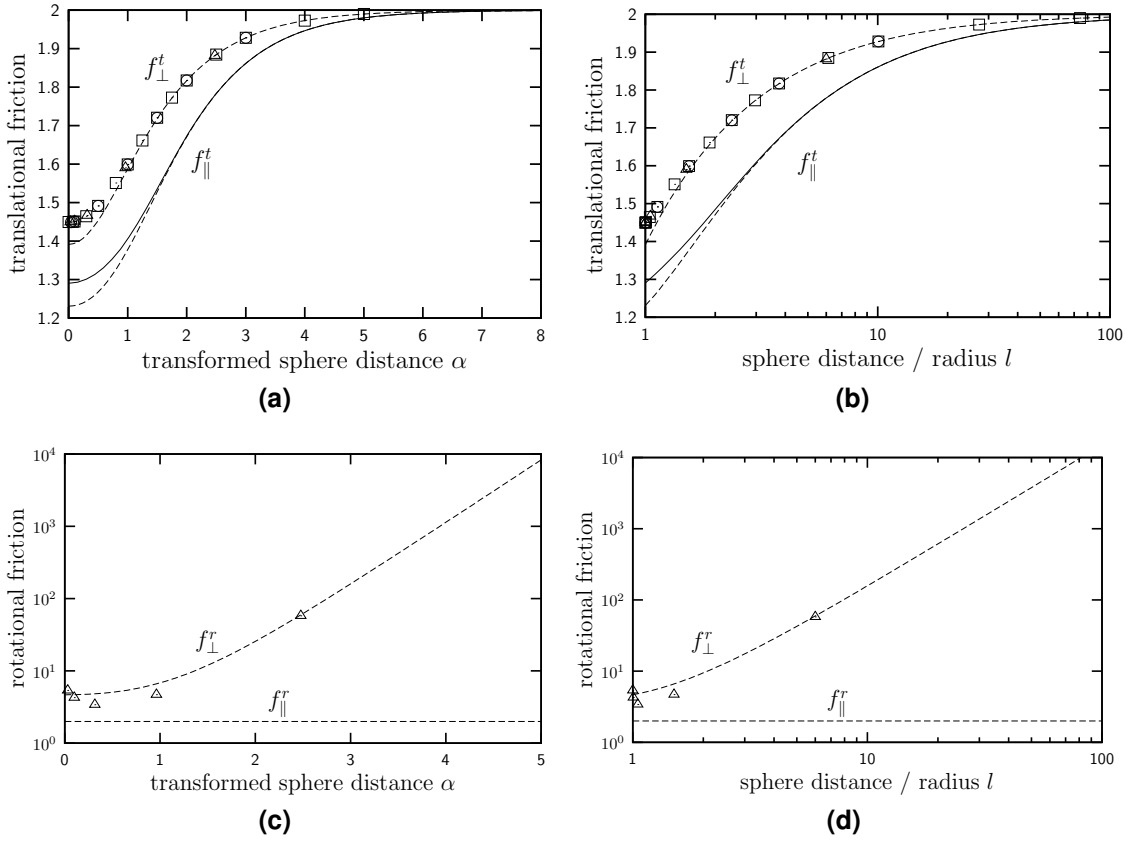


Figure 2.9: Comparison of approximate results from the method explained in the previous section and (quasi-) exact results for the hydrodynamic properties of a rigid dumbbell. All plots show coefficients of translational (a), (b) and rotational (c), (d) friction normalized by the corresponding value for a single sphere. The distance of the two spheres in the dumbbell is given in terms of $\alpha = \operatorname{arcosh} L/R$ in (a), (c) and by $l = L/R$ on a logarithmic scale in (b), (d). The dashed lines denote the approximations by de la Torre et al. as given in Eq. (2.91). The solid line in (a) and (b) is the exact solution for $f^t_\parallel$ according to Ref. [143]. The data points marked by hollow triangles, squares and circles denote numeric results from Refs. [38], [67], and [156] respectively.

also $\alpha = \beta$. In most of the analytic solutions, α is used as parameter to describe the distance of the spheres, which particularly means:

$$\alpha = \ln\left(\frac{L}{R} + \sqrt{\left(\frac{L}{R}\right)^2 - 1}\right) = \operatorname{arcosh}\frac{L}{R}\,. \tag{2.98}$$

The two limits of interest discussed above $l = 1$ and $l \to \infty$ are therefore equivalent to $\alpha = 0$ and $\alpha \to \infty$, respectively. In this coordinate space the differential equations of incompressible viscous hydrodynamics in the time-independent Stokes approximation can be solved under the the boundary condition of two spheres for several types of motions. The only closed form of such a solution was given by Stimson and Jeffery [143]. The authors considered the hydrodynamic friction force acting on one sphere moving parallel to $\mathbf{e}_z$ with constant speed in the presence of another sphere. For the sake of comparability, all results are normalized by the according friction of a single sphere $k_B T_a / D^t_{\mathrm{sph}}$ and $k_B T_a / D^r_{\mathrm{sph}}$.

The work of Stimson et al. [143] gives the following expression:

$$f_{\parallel}^{t} = \frac{k_B T_a}{D_{\mathrm{sph}}^{t}} \frac{4}{3} \sinh\alpha \sum_{n=1}^{\infty} \frac{n(n+1)}{(2n-1)(2n+3)} \left\{ 1 - \frac{4\sinh^2\left(a_n\alpha/2\right) - a_n^2 \sinh^2\alpha}{2\sinh^2\left(a_n\alpha\right) - a_n^2\sinh(2\alpha)} \right\} , \quad (2.99)$$

with $a_n = (2n+1)$. Note that neither of the two limits of small and large separations can be carried out straightforwardly as they involve an infinite sum over vanishing terms leading to finite values. However, the limits can be carried out numerically and one particularly finds:

$$f_{\parallel}^{t} \approx 1.2903 \qquad \Leftrightarrow \qquad f_{\parallel,\,\mathrm{Torre}}^{t} = \frac{16}{13} \approx 1.2308 \, . \quad (2.100)$$

Goldman et al. solved the problem of two spheres of equal size both moving with the same speed perpendicular to $\mathbf{e}_z$ [67]. Wakiya developed a rather general notation for the solution of this class of questions and gave expressions for equal-sized spheres in uniform translational and rotational motion [156]. In Ref. [38], Davis extended this treatment to the even more general problem of two spheres of different size translating with speeds U_1 and U_2 along $\mathbf{e}_1 \perp \mathbf{e}_z$ and rotating with speeds Ω_1 and Ω_2 around $\mathbf{e}_2$, where $\mathbf{e}_1 \perp \mathbf{e}_2 \perp \mathbf{e}_z$. Note that the distance of the spheres is given in terms of the smallest distance between their surfaces S, which would be $S = 2L-2R = 2R(l-1)$ in our case. No closed expressions are available for these cases in contrast to the work of Stimson et al. discussed above. In Refs. [38, 67, 156] the solutions are expressed in terms of infinite sums, while the coefficients must be found by solving a system of linear equations. As the sums as well as the respective coefficients can be proofed to converge, in practice, they are cut off at some finite index N by setting the N^{th} coefficient in the sum $\equiv 0$. This leads to a *finite* system of equations which can be solved numerically. The smaller the distance of the spheres the slower is the convergence. Therefore the limit $\alpha \to 0$ is non-trivial. However, in Ref. [67] the authors give a numerical value for $f_{\perp}^{t}$:

$$f_{\perp}^{t} \approx 1.4494 \qquad \Leftrightarrow \qquad f_{\perp,\,\mathrm{Torre}}^{t} = \frac{32}{23} \approx 1.3913 \, . \quad (2.101)$$

In all papers, results of numerical computations for specific problems are given, which are quasi-exact and can thus be used for comparison with the results of the de la Torre approximation. Fig. 2.9 shows the approximations according to Eq. (2.91) alongside (quasi-) exact results. Obviously, $f_{\parallel}^{t}$ and $f_{\perp}^{t}$ are both well reproduced except for small separations, where the friction is slightly underestimated.

Unfortunately, in the cited papers the torques due to hydrodynamic rotational coupling are only evaluated for freely rotating spheres and not for a rotating dumbbell of spheres which are fixed relative to each other as in the case considered here. However, a reasonable measure for $f_{\perp}^{r}$ can be constructed from the translational friction of the angular motion with the velocity ωL as explained above. As the velocity vectors of the two spheres are antiparallel, only the treatment of Davis [38] is general enough to use it for the respective calculation: $f_{\perp}^{r} = 2(\zeta^{t}(L) - \zeta_{c}^{t}(L))L^2$. The comparison with de la Torre's results is shown in Figs. 2.9c and 2.9d.

Chapter 3

Bimolecular encounter

The elementary unit of a multitude of cellular processes is the bimolecular protein-protein interaction. It plays a key role in many cellular processes such as signal transduction, bioenergetics, and the immune response [77]. The strength and specificity of protein-protein association are mainly determined by the complementarity of protein shape and often by the electrostatics of the protein surfaces. This was shown by experimental measurements at different ionic strengths [126]. Kinetic on-rates are commonly believed to be controlled by long-ranged electrostatic interactions, whereas off-rates are rather controlled by short-ranged interactions like hydrogen bonding and van der Waals forces. Therefore environmental control of complex dynamics has to be implemented through on- rather than off-rates. Notable exceptions are situations in which mechanical force is involved, like in the cytoskeleton.

In this study [124] we are interested in describing how general principles guiding the diffusional association of biomolecular pairs are modulated by their particular physico-chemical properties. We systematically explore the effect of various coarse-graining procedures on the rate of protein-protein encounter. We combine early approaches based on Langevin equations with current knowledge on molecular structure. We apply our models to three systems with different physico-chemical characteristics.

3.1 Biologic examples at different levels of detail

One of the most well-observed bimolecular complexes is the extracellular ribonuclease barnase and its intracellular inhibitor barstar. Both proteins carry a net charge of $2e$ and $-6e$, respectively, which leads to a considerable electrostatic steering [36, 46, 131, 133, 157]. Considering the structure of the two proteins, barnase has a bean-like form, matching well on a large reactive area with the nearly spherical barstar. A classic example of electrostatically-driven protein association is the iso-1-cytochrome *c* - cytochrome c peroxidase (cytc:ccp) complex, charged with $6e$ and $-13e$, respectively and exhibiting dipoles aligned well with the reactive areas [99, 107]. Finally, we selected the medically important complex of a peptide fragment of p53 and its inhibitor mdm2, which is widely used for an-

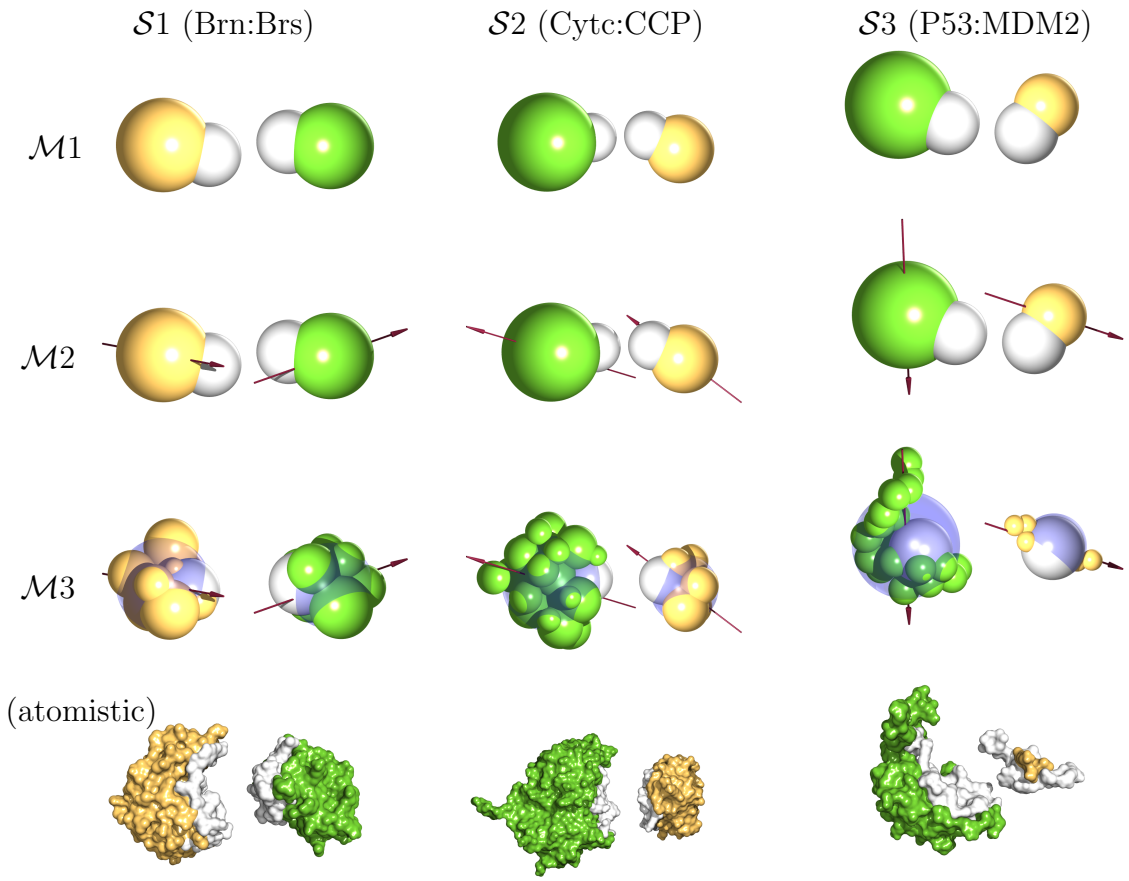


Figure 3.1: Scheme to visualize the different variants of the model for the three considered systems $\mathcal{S}1$–$\mathcal{S}3$. Barnase, cytochrome c and p53 are shown in light gray; barstar, cytochrome c peroxidase and mdm2 are shown in dark gray. The respective reactive patches are shown in white. Model $\mathcal{M}1$ only includes a simple steric interaction. Model $\mathcal{M}2$ has an additional effective electrostatic interaction, here denoted with black arrows showing the direction of the dipole of the model particles. In $\mathcal{M}3$, the excluded volume is modeled in more detail as a collection of smaller beads. The transparent spherical surface marks the volume used in $\mathcal{M}1$ and $\mathcal{M}2$ for the sake of comparison. Finally, the bottom row shows surface representations of the atomistic structures taken from the protein data bank.

ticancer drug design. In this system, electrostatic attraction plays a very minor role. On the other hand, the steric matching of both surfaces is of particular importance here. It is a perfect example of a key-lock binding interface, where p53 is buried deep into a cleft on the mdm2 surface.

One aim of this chapter is to determine how crucial particular details of the model proteins are with respect to the association properties. Therefore, we consider a system of model proteins at three different levels of detail as depicted in Fig. 3.1. We treat it with the numerical Langevin equation approach already introduced in chapter 2. In the most generic approach ($\mathcal{M}1$), we only account for the steric interaction between spherical particles. As a first refinement ($\mathcal{M}2$), an effective Coulombic interaction is introduced using the dipolar sphere model

(DSM – see Sect. 2.1). Finally, since our Langevin equation approach is particularly suited to capture anisotropic transport, we implement a more refined version for protein sterics ($\mathcal{M}3$). In this approach the excluded volume of each protein is modeled by 8–25 smaller beads. $\mathcal{M}3$ uses the DSM as well.

The simulations were performed in a cubic box with periodic boundary conditions. For the application of the DSM this particularly means that actually an infinite number of copies exists for each model charge. However, due to the very quick decay of the electrostatic interaction, only the *minimum image distance* of two charges is considered in the force calculation. That is, each charge interacts with only the closest copy of any other charge in the system due to the periodic boundary conditions.

Schreiber and Fersht used concentrations between $0.125\mu\mathrm{M}$ and $0.5\mu\mathrm{M}$ in their experimental studies of the association rate of the barnase:barstar complex [126]. The average volume containing one particle at a concentration c is $1/cN_A$ with the Avogadro number $N_A = 6 \cdot 10^{23}\mathrm{mol}^{-1}$. Hence, the edge length of a cubic boundary box representing concentration c can be calculated from:

$$L = \sqrt[3]{V} = 1/\sqrt[3]{cN_A}\,. \tag{3.1}$$

For example, $c = 0.125\mu\mathrm{M}$ leads to $L \approx 2370\text{Å}$ for one pair of particles, which is two orders of magnitude larger than the size of the proteins. Due to this low density, the first passage times (FPT) to encounter can be expected to be much longer than the chosen time step. For this reason we use the variable time step approach in our simulations as it has been explained in chapter 2.

Gabdoulline and Wade [58] used several criteria to define a contact area for the system. In our studies, we define the contact area to consist of those atoms in the two interacting proteins that are at 5Å or less distance from an atom of the complementary protein. The center of mass of these atoms is considered as the center of the reactive area. For $\mathcal{M}1$ and $\mathcal{M}2$, the reactive patch is centered at the surface of the sphere modeling the excluded volume such that it has the same relative direction from the center of mass as obtained by the method just described. In the case of $\mathcal{M}3$, the center of the patch is set to the center of the reactive area.

The contact area has a diameter of approximately 10Å to 20Å for the three systems studied here. Following earlier Brownian Dynamics simulations [140] Alamanova et al. performed an in-depth analysis of the free energy landscape and the encounter state of the protein complexes considered in this chapter [3]. Both studies showed that the encounter complex is typically located at relative separations of the two protein surfaces of about 10Å compared to their positions in the final complex. As the reactive patches in this study model both the size of the contact area on the surface of the model particles and the distance above their surface at which an encounter will be possible, values in the range of 5Å to 10Å seem reasonable.

Two types of excluded volume structures are taken into account. In the first case, used in $\mathcal{M}1$ and $\mathcal{M}2$, the proteins are assumed to have an approximately spherical form. The radius for the model spheres determining the hard core interaction is determined by the radius of gyration of the protein. This as well as the

Table 3.1: Protein structures and parameters used in the study. The last column gives the angle between the position of the patch relative to the center of mass ($\mathbf{r}_{\mathrm{patch}}$) and the dipole moment ($\mathbf{p}$). Note that two corresponding patch vectors $\mathbf{r}_{\mathrm{patch}}$ are usually close to antiparallel, while two corresponding dipole moments $\mathbf{p}$ electrostatically favour a parallel alignment. That is, the most beneficial configuration would be if $\angle_1 = 0°$ and $\angle_2 = 180°$.

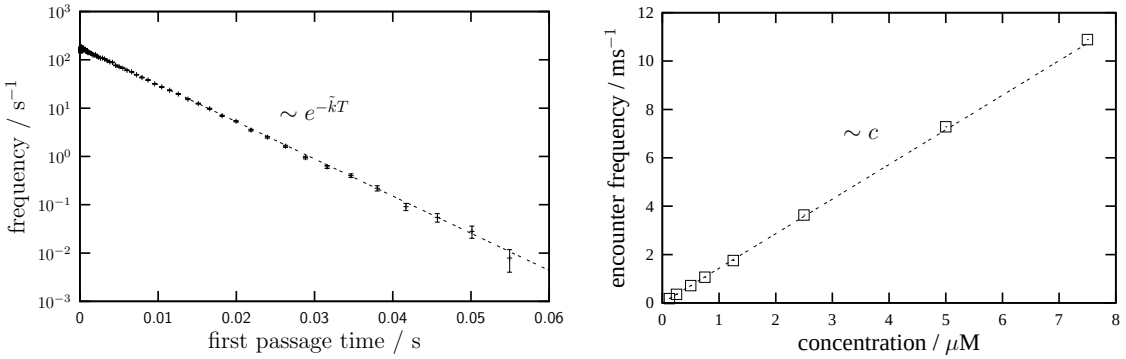
Protein	System	PDB code	Ref.	$R_{\mathrm{gyr}}/\mathrm{Å}$	q/e	$\angle(\mathbf{r}_{\mathrm{patch}}, \mathbf{p})$
barnase	$\mathcal{S}1$	1BRS	[30, 58]	14.68	2	59.5°
barstar				13.42	−6	135°
cytc	$\mathcal{S}2$	2PCC	[107]	13.89	6	22.3°
ccp				20.00	−13	153.8°
p53	$\mathcal{S}3$	1YCR	[88]	10.20	−2	110°
mdm2				16.81	1	83°

anisotropic diffusion matrix of the proteins is calculated by the HYDROPRO software [40]. The underlying data in the more detailed approach $\mathcal{M}3$ is obtained using the AtoB bead modeling software [31, 114]. In this way, the three-dimensional structure of the proteins is modeled with a comparably small number of 8 to 25 spheres of different sizes.

The charges for the DSM were calculated as explained in Sect. 2.1 according to the atomistic structures obtained from the protein data bank (PDB) [24]. The data are summarized in Tab. 3.1 for the proteins we consider in our studies.

3.2 General scaling behavior

Langevin dynamics simulations were performed for cubic boxes containing two model proteins. Simulations were conducted until the encounter condition was met for the first time over typically millisecond lengths. The central quantity to be determined in the simulations is the *encounter frequency* $\tilde{k}$ for a given parameterization of the model. Instead, the accessible quantity in our case is the first passage time (FPT) T to encounter. For a Poisson-like process, where the distribution of FPT is given by $f(T) = \tilde{k}e^{-\tilde{k}T}$, the encounter frequency is the reciprocal of the mean FPT $\tilde{k} = 1/\langle T\rangle$. Later we will refer to the encounter frequency scaled by the concentration as the *encounter rate* $k = \tilde{k}/c$. It is reasonable to refer to the unscaled quantity $\tilde{k} = 1/\langle T\rangle$ as frequency as it has the dimension s^{-1}. As the preparation of a comparable experiment would never allow knowing the particular initial positions and orientations of the unbound proteins, it makes sense to average over the possible initial configurations. Therefore, we started a large number of runs (typically 10^4 to 10^5) with random initial positions and orientations for all



(a) First passage time histogram

(b) Encounter frequency for several concentrations

Figure 3.2: (a) Logarithmic plot of the distribution of the FPT to encounter T between a single pair of barnase and barstar model particles in a cubic boundary box of edge length $L = 2370\text{Å}$, representing a concentration of $0.125\mu\text{M}$ for each protein. The dashed line represents a single exponential fit to the data points, which shows the expected behavior with respect to the encounter frequency $\tilde{k} = \langle T \rangle^{-1}$. (b) Simulated encounter frequencies for a single pair of barnase and barstar model particles in cubic boundary boxes of different sizes representing different concentrations. The dashed line is a linear fit to the data.

involved model particles, under the constraint that the initial pairwise distance is at least large enough to prevent an immediate encounter. The "first passage" is defined as the first overlap of two complementary reactive patches. Interestingly, by this averaging, the first passage process becomes Poisson-like, see Fig. 3.2a. The data show a clear exponential behavior. This means that it is truly reasonable to talk of an "encounter rate", as the FPT distribution is indeed reflected by a single stochastic rate. The finite probability at small FPT is due to the possibility that the two model particles are initialized in close proximity in the beginning of the simulation. The large errors in the histogram at $T \to 0$ are caused by the fact that exponentially sized histogram bins were used to sample the behavior for small T. Therefore, events hitting a particular bin are rare because of the small width of the bins at $T \to 0$, which then leads to bad statistics in this domain.

As the encounter process is purely diffusion limited, one expects the encounter frequency to scale linearly with the concentration. The plausible argument for this is the rising number of combinatorial possibilities per volume with decreasing concentration. Fig. 3.2b shows the encounter frequencies for the barnase:barstar model system without electrostatic interaction ($\mathcal{M}1$). The linear behavior is well reproduced. Hence, it is reasonable to always scale the encounter frequencies with the inverse concentration for a better comparability, as it will be done for the remaining part of this work. We denote the rescaled quantities as *encounter rate* in the following, i.e. the encounter rates have the dimension $\text{M}^{-1}\text{s}^{-1}$.

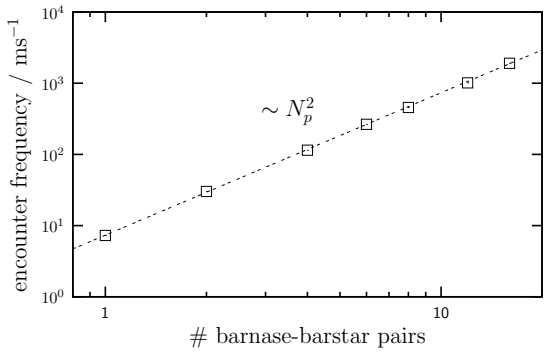


Figure 3.3: Encounter frequency for different numbers of barnase:barstar pairs leaving the size of the boundary box constant. The data points show the encounter frequencies as obtained from simulations, while the solid line represents the function CN_p^2, where N_p is the number of particle pairs and C is a prefactor obtained by a fit to the data.

Finite size effects

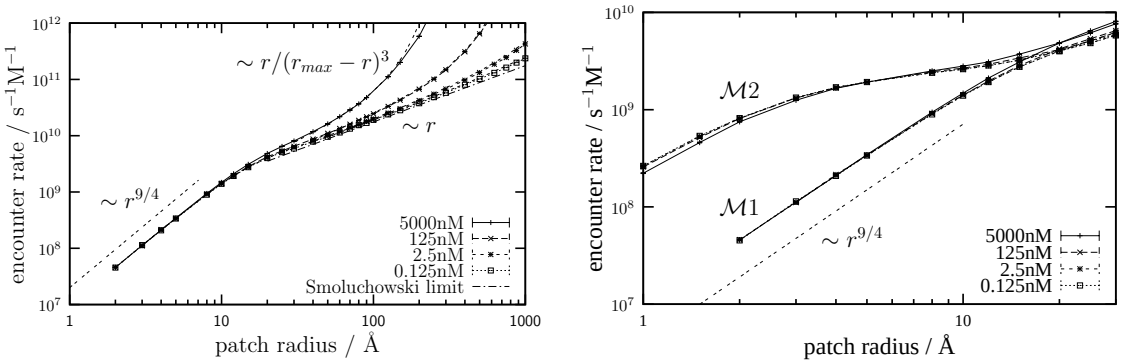
In most of the simulations, only one instance of the final complex was considered, i.e. one model particle of each kind. Using such small systems could lead to undesired finite size effects. Fig. 3.3 shows the simulation results for k^{enc} for an increasing number of barnase:barstar pairs, while keeping the size of the boundary box constant. The idea is to rule out that larger numbers of particles may lead to noticeable three-body interactions or hindering of the encounter process. In particular, this means that just the number of possible pairs for binding increases. Consider a system of two barnase molecules (A and A') and two barstar molecules (B and B') randomly distributed over the boundary box. The relative alignment of any pair of As and Bs is therefore random again. For a particular pair the distribution of times to first encounter will thus look very similar to the case with a single pair in the box, which is a simple exponential decay with respect to the encounter frequency k_1^{enc}: $f_1(T) = k_1^{\mathrm{enc}} \exp[-k_1^{\mathrm{enc}} T]$. The probability that, e.g., the particular pair $\mathrm{A} - \mathrm{B}$ reaches encounter at a certain time t before the three other possible pairs ($\mathrm{A}' - \mathrm{B}$, $\mathrm{A} - \mathrm{B}'$, $\mathrm{A}' - \mathrm{B}'$), is therefore:

$$\begin{aligned} p(t) &= \int_0^\infty \mathrm{d}t_1 \int_{t_1}^\infty \mathrm{d}t_2 \int_{t_1}^\infty \mathrm{d}t_3 \int_{t_1}^\infty \mathrm{d}t_4 \, \delta(t_1 - t) \prod_{i=1}^{4} k_1^{\mathrm{enc}} e^{-k_1^{\mathrm{enc}} t_i} \\ &= k_1^{\mathrm{enc}} e^{-4 k_1^{\mathrm{enc}} t} \; . \end{aligned} \tag{3.2}$$

Thus, the probability that any of the four possible particle pairs reaches encounter before the respective three other pairs do, is $4 \times p(t)$ as just calculated, i.e. $f_2(T)$ has again a Poisson form like $f_1(T)$ and $k_2^{\mathrm{enc}} = 4 k_1^{\mathrm{enc}}$. In general, for higher numbers of particle pairs N_p, it is expected to again find an exponential distribution of the time to first encounter with the encounter frequency $k_{N_p}^{\mathrm{enc}} = N_p^2 k_1^{\mathrm{enc}}$. This quadratic behavior is again well reproduced by the data shown in Fig. 3.3, which suggests that even for small systems with only two particles no strong finite size effects have to be expected.

Size of the reactive patches

We want to understand the dependence of the data on the size of the reactive patches in more detail. This behavior is exemplarily studied with the bar-



(a) Influence of patch radius **(b)** Including effective electrostatics

Figure 3.4: (a) Encounter rates in dependency of the patch size for the barnase:barstar model system in the $\mathcal{M}1$ variant. (b) Comparison of $\mathcal{M}1$ and $\mathcal{M}2$ similar to Fig. 3.4a for small patch sizes. For larger patch sizes there is no substantial difference.

nase:barstar model system. In Fig. 3.4a, the encounter frequency has been obtained from simulations for barnase:barstar-like model particles in the framework of $\mathcal{M}1$ at several concentrations $c = \{5\mu\mathrm{M}, 125\mathrm{nM}, 2.5\mathrm{nM}, 125\mathrm{pM}\}$ and varying patch sizes r. All values in the figure have been scaled with the concentration, which leads to a collapse as expected. It is obvious that as r gets larger than $2R$ at around $r = 40\text{Å}$, the reactive patch "covers" the whole model particle and we therefore go over to the Smoluchowski limit of isotropic reactivity, where $k_{\mathrm{smo}} \sim r$. This gets particularly apparent in the chain dotted line, which depicts the Smoluchowski approximation according to Eq. (1.1) with $D = D_{\mathrm{Brn}} + D_{\mathrm{Brs}} \approx 2.33 \cdot 10^{-10}\mathrm{m}^2/\mathrm{s}$ chosen as the combined translational diffusion coefficient of barnase and barstar and assuming the capture radius is determined by the patch size $r_0 = r$, which is true for large r. The simulation results are in good agreement with the Smoluchowski approximation in an intermediate range of r. However, at high densities or large r respectively, the patches span a serious part of the simulation box of edge length L, and do immediately encounter for a threshold value of $r = r_{max} = L\sqrt{3}/4$, where the sum of the patch diameters $4r$ equals the triagonal. Thus, the encounter frequency must diverge with $\sim 1/(r_{max} - r)^{\alpha}$, where we suppose $\alpha = 3$, as the volume of configurational space without immediate encounter is decreasing with r^3. This assumption in addition with the Smoluchowski behavior would lead to $k \sim r/(r_{max} - r)^3$ for large r, which follows the data in Fig. 3.4a (dashed lines).

On the other hand, the scaling for smaller r is clearly not $k \sim r^4$, as one would expect from naive geometric arguments. This is consistent with the findings of Shoup et al. [135], who stated that the association rate of similar systems with spherical reactive patches is considerably increased due to the effect of rotational diffusion. Furthermore, if two complementary particles come into close proximity, a process as described by Smoluchowski, they will stay close for a finite time and test not only one orientation but part of the orientational configuration space, until they completely disperse again, as e.g. stated in Ref. [100]. Therefore, the

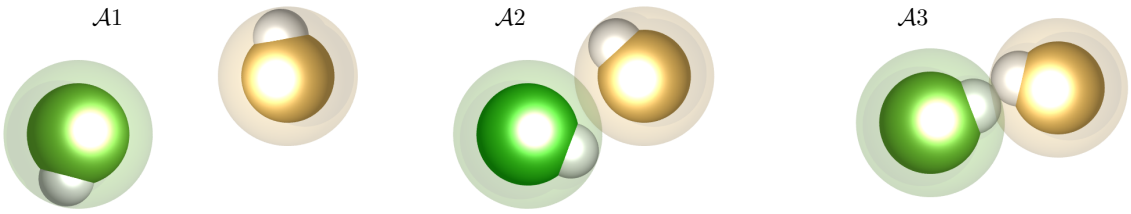


Figure 3.5: Different alignment states during the encounter process. $\mathcal{A}1$ proteins are completely unaligned. In state $\mathcal{A}2$, referred to as *contact* in this chapter, the proteins are translationally aligned, i.e. they are close enough to actually encounter (denoted by the overlap of the lightened area around the model particles), but lack the correct orientation. $\mathcal{A}3$ proteins reached the encounter meaning that the reactive patches are in translational and rotational alignment.

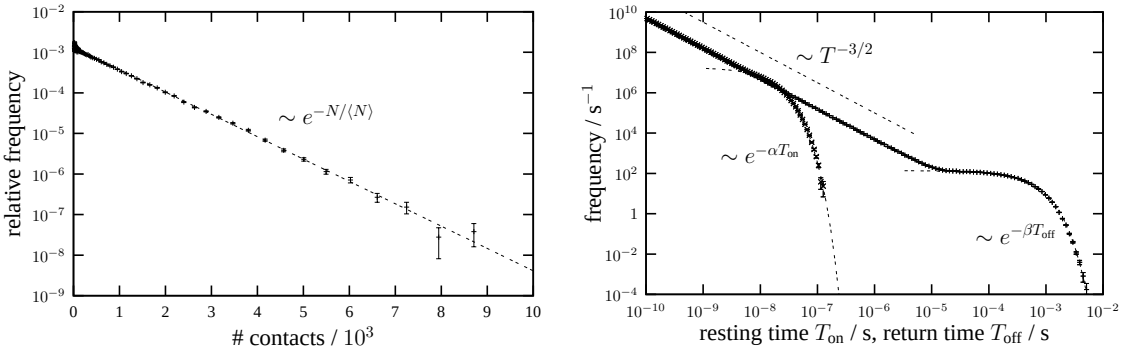
question if the encounter is reached when two particles are at a small distance is not answered by a static geometric probability.

It has been shown that the electrostatic interaction of proteins can strongly increase the association rate. However, under physiological salt conditions, Coulombic interactions are screened by counter ions in the solution on a small length scale of approximately $\kappa = 1\mathrm{nm}$. Thus, deviations from case $\mathcal{M}1$ without effective charges will only arise for small r. Fig. 3.4b shows the results of respective simulations of our model $\mathcal{M}2$ compared to the results of $\mathcal{M}1$. Indeed, for large patch radii r, the results are similar, while for smaller r, the encounter rates in $\mathcal{M}2$ are clearly higher compared to $\mathcal{M}1$. However, the crossover to a power law behavior with roughly $\sim r^{9/4}$ can be suspected for very small r, just at a prefactor of about 50 times larger than for $\mathcal{M}1$.

3.3 Study of the encounter pathway

The pathway through which the encounter is formed is of special interest. We dissect the encounter process into several states as illustrated in Fig. 3.5. At the start of each run, the systems were prepared in the unaligned state $\mathcal{A}1$, as described earlier. We assume that the two model proteins will then switch between state $\mathcal{A}1$ and $\mathcal{A}2$ a number of times N, until they finally reach the encounter complex $\mathcal{A}3$ due to rotational diffusion. In the following, each of these switches from $\mathcal{A}1$ to $\mathcal{A}2$ will be termed a *contact*. That is, N counts the number of unsuccessful contacts before the encounter is finally formed. A separate set of simulations was performed to measure the distribution of N. Furthermore, we analyzed the distribution of return times T_{off}. This is the time it takes for two model proteins to get into contact again after having lost translational alignment ($\mathcal{A}2 \rightarrow \mathcal{A}1$), i.e., they have been in close proximity once. Finally, we determined the distribution of resting times T_{on} in translational alignment $\mathcal{A}2$ before the two model particles separate again.

As an example, Fig. 3.6a shows the distribution of N for the barnase:barstar model system at $c = 0.5\mu\mathrm{M}$ in the framework of $\mathcal{M}1$. The distribution of the number of contacts again has a Poisson form. This is not obvious. Although,



(a) Histogram of number of contacts N **(b)** Resting and return time histogram

Figure 3.6: (a) Logarithmically plotted distribution of the number of contacts N between a barnase and a barstar particle. The dashed line is an exponential fit to the data. (b) Double-logarithmic plot of the distribution of resting and return times of the translationally aligned state ($\mathcal{A}2$ in Fig. 3.5).

as we start with random initial conditions, low numbers of approaches until the encounter patches overlap can be expected to occur with a finite probability. Interestingly, the number of unsuccessful contacts in state $\mathcal{A}2$ can be rather large. We also find that the distribution of N is roughly independent of the concentration. This is reasonable, as after the two proteins were in contact once, the further encounter process is guided by returns to state $\mathcal{A}2$ and thus is virtually independent of the system size.

However, Fig. 3.6b shows that the return time $T_{\rm off}$ is not exponentially distributed. Instead, it follows a power law $p(T_{\rm off}) \sim T_{\rm off}^{-3/2}$ and undergoes an exponential cutoff due to the finite size of the boundary box at large $T_{\rm off}$. Therefore, there is a high probability for very small return times, i.e. situations, where the two model proteins do not really disperse, but immediately after loosing translational alignment ($\mathcal{A}2 \rightarrow \mathcal{A}1$) get into contact again ($\mathcal{A}1 \rightarrow \mathcal{A}2$). The power law behavior of the return time is consistent with the problem of a random walk to an absorber in three dimensions [112]. In principle, these two situations are equivalent since the relative motion of the two proteins while unaligned ($\mathcal{A}1$) can be approximately understood as a random walk, and the criterion for going over to translational alignment $\mathcal{A}2$ reflects an absorbing boundary in the configuration space of relative positions. The distribution of resting times $T_{\rm on}$ follows the same power law as $f(T_{\rm off})$, but the exponential cutoff happens much earlier. The reason is that the finiteness is here determined by the region in configuration space, where the two model proteins are in state $\mathcal{A}2$. As this is much smaller than the whole volume of the boundary box, in which they are unaligned and therefore in state $\mathcal{A}1$, a random walk in state $\mathcal{A}2$ will end much earlier. The power law in both histograms for $T_{\rm on}$ and $T_{\rm off}$ makes it actually impossible to consistently normalize them because of the divergence for $T \rightarrow 0$. However, a convenient normalization can be achieved by introducing some cut-off $T_{\rm min}$. This is technically reasonable as we use a numerical scheme for the evolution of the system with a discrete and finite minimal time step $\Delta t_{\rm min}$. And also physically it makes sense to cut off the histogram as

the applied concepts of diffusion break down at very small time scales when motion gets ballistic and therefore the picture would change at $T \to 0$ systematic reasons.

The differences we obtain in the distributions of T_{on} and T_{off} when using the variants $\mathcal{M}2$ and $\mathcal{M}3$ compared to $\mathcal{M}1$ are generally very small and unlikely to account for any deviations in the overall encounter rates. Also, the distribution of N is always well described by a single exponential decay. However, the inverse decay length $\langle N \rangle$ really varies for the different situations. Therefore, changes in the overall encounter rate are mainly caused by a different probability for reaching state $\mathcal{A}3$ being in contact ($\mathcal{A}2$). This is reasonable when considering that all regarded effects are strongly localized and can thus only act while the system is in contact.

3.3.1 Three systems with different physico-chemical interface properties

So far we have exclusively considered a parameterization of barnase:barstar ($\mathcal{S}1$) to demonstrate the principal behavior of our model and how we access particular quantities. As the techniques we use for finding parameter values are general, we now apply our method to two other systems cytochrome c and its peroxidase ($\mathcal{S}2$) as well as the p53:mdm2 complex ($\mathcal{S}3$). Those represent systems with different interface characteristics and where the role of electrostatics is either much stronger ($\mathcal{S}2$) or much weaker ($\mathcal{S}3$) than for $\mathcal{S}1$.

Basically all the previously described quantities are measured for 8 different concentrations $c = \{125, 250, 500, 750, 1250, 2500, 5000, 7500\}\mathrm{pM}$. Furthermore, to find out how crucial the choice of the radius of the reactive patch affects the results, we use patch radii of $r = 6\text{Å}$ and $r = 3\text{Å}$ in addition to the initially considered value of $r = 10\text{Å}$.

Tab. 3.2 shows the encounter rates k and mean number of contacts $\langle N \rangle$ as obtained from the simulations. The rates are all roughly of the same order of magnitude. Still, several qualitative features are readily apparent. First, for decreasing patch sizes, the rates generally decrease as expected. Secondly, this effect is weaker for $\mathcal{M}2$ compared to $\mathcal{M}1$, which basically means that the electrostatic attraction and orientation due to the dipole interaction are indeed enhancing the encounter. The strongest effect of the electrostatic interaction is obtained for cytc:ccp, which is the system with the largest monopole and dipole and the best alignment of the directions of the dipoles and the reactive patches. On the other hand p53:mdm2 is nearly unaffected by the effective charges, due to its weak monopole charges and, additionally, a rather bad alignment of the dipolar interaction and the reactive surface area. Furthermore, regarding the results with detailed steric structure $\mathcal{M}3$, the effect on the rate is correlated with the deviations of the protein forms from the spherical excluded volume approach in $\mathcal{M}1$ and $\mathcal{M}2$. This deviation is smallest for cytc:ccp and largest for p53:mdm2.

The findings for the encounter rate k are also reflected in the results for $\langle N \rangle$. As expected, higher $\langle N \rangle$ leads to a decrease in k. This principle is very con-

Table 3.2: All values $k(\mathcal{M}\mathrm{x})$ denote encounter rates which have been averaged over several simulations at different concentrations as given in the text. The values are given in k / $10^9\mathrm{M}^{-1}\mathrm{s}^{-1}$ for the three different versions of our model. $\langle N\rangle(\mathcal{M}\mathrm{x})$ are average values for the mean number of contacts. As already mentioned, $\langle N\rangle$ is basically independent of the concentration. Therefore we average over the different simulations for each of the chosen systems. The errors were determined by one standard deviation from the 8 values obtained at different concentrations. Some of the choices for the patch radius were not applicable to $\mathcal{M}3$, as for these cases an encounter was completely hindered by the detailed excluded volume model.

System	Patch radius	$k(\mathcal{M}1)$	$k(\mathcal{M}2)$	$k(\mathcal{M}3)$
brn:brs	10.0	1.56 ± 0.04	2.76 ± 0.07	2.02 ± 0.02
	6.0	0.57 ± 0.01	2.13 ± 0.01	1.34 ± 0.08
	3.0	0.13 ± 0.001	1.28 ± 0.03	$-$
cytc:ccp	10.0	1.12 ± 0.02	4.31 ± 0.20	4.15 ± 0.15
	6.0	0.40 ± 0.01	4.29 ± 0.21	4.05 ± 0.09
	3.0	0.09 ± 0.001	4.03 ± 0.05	0.21 ± 0.02
p53:mdm2	10.0	2.05 ± 0.05	2.51 ± 0.04	1.27 ± 0.02
	6.0	0.80 ± 0.01	1.12 ± 0.01	0.15 ± 0.01
	3.0	0.19 ± 0.002	0.28 ± 0.01	$-$
System	Patch radius	$\langle N\rangle(\mathcal{M}1)$	$\langle N\rangle(\mathcal{M}2)$	$\langle N\rangle(\mathcal{M}3)$
brn:brs	10.0	474 ± 2	198 ± 5	282 ± 10
	6.0	1140 ± 4	232 ± 8	534 ± 50
	3.0	4120 ± 10	653 ± 15	$-$
cytc:ccp	10.0	842 ± 3	61 ± 5	71 ± 7
	6.0	2040 ± 20	30 ± 3	77 ± 10
	3.0	7540 ± 63	21 ± 3	4160 ± 375
p53:mdm2	10.0	362 ± 2	266 ± 4	823 ± 10
	6.0	815 ± 10	582 ± 13	8720 ± 200
	3.0	2900 ± 35	2550 ± 30	$-$

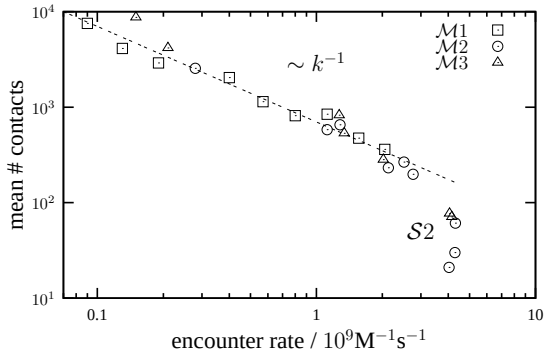


Figure 3.7: Correlation plot of encounter rate k and mean number of contacts $\langle N \rangle$ including all data from Tab. 3.2.

sistent with the data. The only exception is cytc:ccp observed in $\mathcal{M}2$. Here, the effective Coulombic interaction is strongest and the dipole moment is best aligned with the reactive patches. Therefore, having reached state $\mathcal{A}2$ once, the proteins do systematically orient towards $\mathcal{A}3$, while they are additionally strongly steered back towards $\mathcal{A}2$ when loosing their translational alignment. This behavior is the stronger the closer the model proteins have approached once – i.e. for the case of small patch sizes, where state $\mathcal{A}2$ implies the smallest distance. While this only explains the inversion in the $\langle N \rangle$ behavior, k is obviously still slightly decreasing with smaller patch sizes. This can be explained by the fact that the time to the first approach of state $\mathcal{A}2$ is larger for smaller patches, as this implies a smaller relative distance. This obviously compensates the fact that afterwards the encounter is formed even quicker, as reflected by the decreasing $\langle N \rangle$.

The strong correlation between the encounter rate k and the mean number of contacts $\langle N \rangle$ is most apparent in the respective correlation plot in Fig. 3.7. Indeed, $k \sim \langle N \rangle^{-1}$ is valid for most of the different systems and models. Note, particularly, that the prefactor is very similar in all cases. Basically, this means that one unsuccessful contact takes the same amount of time on average, no matter what the local details of the system are. This gets more obvious recalling the distributions of the resting and return times T_{on} and T_{off} in Fig. 3.6b, where it gets clear that $\langle T_{\mathrm{off}} \rangle > \langle T_{\mathrm{on}} \rangle$. As the average time for one contact will be approximately $\langle T_{\mathrm{on}} \rangle + \langle T_{\mathrm{off}} \rangle$, it is dominated by T_{off}, which is only marginally influenced by the local details of the system and the chosen model. Therefore it can be concluded, that for $\mathcal{S}1$ and $\mathcal{S}3$ the incorporation of a more detailed modeling approach influences k and $\langle N \rangle$, but not the overall characteristics of the encounter process.

The only exceptions for the clear correlation of k and $\langle N \rangle$ are $\mathcal{M}2$ and $\mathcal{M}3$ with the cytc:ccp parameterization ($\mathcal{S}2$), where k is nearly independent of $\langle N \rangle$ because of the strong electrostatic interaction. This is consistent with the earlier finding that the behavior of cytc:ccp is qualitatively different [99], as its electrostatic interactions would 'facilitate long-lived nonspecific encounters between the proteins that allowed the severe orientational criteria for reaction to be overcome by rotational diffusion'. In $\mathcal{M}3$, the smallest patch size $r = 3\text{Å}$ leads to an artificial slow-down, because in this case an overlap of the patches is rather hindered by the beads modeling the protein structure.

3.3.2 Analytic approximations

Recalling the *contact* situation as illustrated in Fig. 3.5, the problem of resting and return times $T_{\mathrm{on}}/T_{\mathrm{off}}$ can be simplified to a random walk in three dimensions with specific boundary conditions. The question whether or not the proteins are in contact can be answered without knowledge of the orientation of the particles. That is, the only relevant information is the relative position $\mathbf{r}$. For the analytic solution we choose the easiest model $\mathcal{M}1$, i.e., particularly without electrostatic interaction. Neglecting the anisotropy of diffusion the evolution of $\mathbf{r}$ becomes a simple random walk with the diffusion coefficient $D = D_1^{\mathrm{t}} + D_2^{\mathrm{t}}$, where $D_{1/2}^{\mathrm{t}}$ denote the translational diffusion coefficients of either protein. Fig. 3.8 illustrates the principle idea of this approach and how the contact can be mathematically defined therein. The system is spherically symmetric. Therefore, we reduce the relative distance vector $\mathbf{r}$ to its absolute value r in the following. The contact radius is denoted with a and is given by $a = R_1 + R_2 + r_1 + r_2$, where $R_{1/2}$ and $r_{1/2}$ are the radii of proteins and patches, respectively. We now follow the mathematical treatment of first passage time problems described in Sect. 2.2. In this framework, the problem of resting and return times imply to start the (relative) random walk close to the contact radius at some distance $r(t=0) = r_0 = a \pm \varepsilon$ with $\varepsilon \ll 1$. This corresponds to a situation where the proteins have just lost contact $r_0 = a + \varepsilon$ for T_{off} or came into contact $r_0 = a - \varepsilon$ for T_{on}. The *passage* is then defined by $r(t) = a$ in either case, i.e., the next contact for T_{off} and the loss of contact for T_{on}. That is, there is an absorbing boundary at $r(t) = a$. Treating the return time to contact as it has been measured in the simulations would additionally imply an infinite number of copies of the absorbing spherical shell at $r = a$ aligned in form of a cubic lattice due to the periodic boundary conditions. Unfortunately, such boundary conditions would be extremely complicated to parameterize. Thus we will only consider the return time to a single absorber in an otherwise free space neglecting the periodic boundary conditions. For the resting time, the random walk is performed inside the spherical shell at $r < a$. On the other hand, there is still the effective spherical steric interaction of the proteins which prohibits distances of $r < R_1 + R_2 = b$. Therefore, in this case a reflecting boundary has to be introduced at $r = b$. In the following, the first passage time distribution is calculated and discussed for these two cases.

Return time to translational encounter

From now on, we simply write t in the calculation instead of T_{off}, especially since t here has a different meaning as T_{off} used in the simulation results as explained above. First, a partial differential equation of the form of the heat equation gives the probability density of finding the particles at distance r at time t:

$$\frac{\partial p(\mathbf{r},t)}{\partial t} = \frac{D}{2} \triangle p(\mathbf{r},t) \ , \tag{3.3}$$

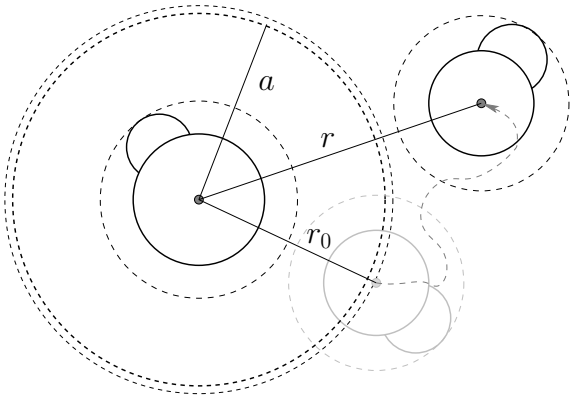


Figure 3.8: The problem of return to contact in a simplified view that can be treated analytically.

which can be rewritten using the spherical symmetry:

$$\frac{\partial p(r,t)}{\partial t} = \frac{D}{2}\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}p(r,t)\ . \tag{3.4}$$

As discussed above in this case, the only boundary condition is an absorbing sphere at $r = a$, where therefore $p(a,t) = 0\,\forall t$. The initial condition is to start at $r = r_0$, which means $p(r,0) = \delta(r - r_0)$. The solution of the above problem with these conditions can be found in Ref. [34]:

$$\begin{aligned} p(r,t) &= \frac{r_0}{r\sqrt{2\pi Dt}}\left\{\exp\left[-\frac{(r-r_0)^2}{2Dt}\right] - \exp\left[-\frac{(r+r_0-2a)^2}{2Dt}\right]\right\} \\ &= \frac{r_0}{r\sqrt{2\pi Dt}}\exp\left[-\frac{(r-r_0)^2}{2Dt}\right]\left\{1 - \exp\left[-\frac{2(r_0-a)(r-a)}{Dt}\right]\right\}\ . \end{aligned} \tag{3.5}$$

The first passage time distribution $f(t)$ is given by the loss of probability density at time t, where "loss" means the negative time derivative. The total probability of not having reached the absorbing boundary $r = a$ yet is the integral of the probability density over the whole space $r \in [a, \infty]$:

$$f(t) = -\int_a^\infty \mathrm{d}r'\,\dot{p}(r',t)\ . \tag{3.6}$$

With $\varepsilon = r_0 - a$, the time derivative of the probability density according to Eq. 3.5 writes:

$$\begin{aligned} \dot{p}(r,t) = \frac{r_0}{\sqrt{8\pi D^3}}&\frac{\exp\left[-\dfrac{(r-r_0)^2}{2Dt}\right]}{r\sqrt{t^5}}\times \\ &\left\{(r-r_0)^2 - Dt - \exp\left[-\frac{2\varepsilon(r+\varepsilon-r_0)}{Dt}\right]\left((r+2\varepsilon-r_0)^2 - Dt\right)\right\}\ . \end{aligned} \tag{3.7}$$

Since we want the initial distance r_0 to be above but close the absorber $r_0 - a = \varepsilon \ll 1$, the term in braces can be expanded around $\varepsilon = 0$:

$$\dot{p}(r,t) = \frac{r_0}{\sqrt{8\pi D^3}} \frac{\exp\left[-\dfrac{(r-r_0)^2}{2Dt}\right]}{r\sqrt{t^5}} (r-r_0)\varepsilon \left(\frac{2(r-r_0)^2}{Dt} - 6\right) + \mathcal{O}(\varepsilon^2) \,. \tag{3.8}$$

The integration over r is still tricky, since generally integration over a Gaussian function only gives a closed solution for special intervals like $\int_0^\infty$, and the integral of the form $\int \mathrm{d}r'\, e^{-r'^2}/r'$ does not have a closed solution at all. The first problem can be tackled by shifting the integration boundaries from $[a,\infty)$ to $[r_0,\infty)$. As $r_0 - a = \varepsilon$ is intended to be small and the relevant contributions around the maximum of the function to integrate will be located at $r > r_0$ for large t also, this is a reasonable approximation and gives solutions for all contributions of the form:

$$\int_0^\infty \mathrm{d}r'\, e^{-cr'^2} (r'+r_0)^\alpha = \begin{cases} \frac{1}{2c}\sqrt{c\pi} & \text{for } \alpha = 0 \\ \frac{1}{2c}\left(1 + r_0\sqrt{c\pi}\right) & \text{for } \alpha = 1 \\ \frac{1}{4c^{3/2}}\left(\sqrt{\pi} + 4r_0\sqrt{c} + 2cr_0^2\sqrt{\pi}\right) & \text{for } \alpha = 2 \end{cases} , \tag{3.9}$$

where r' is the shifted $r - r_0$ and therefore $r \to r' + r_0$. As already stated, the case of $\alpha = -1$ is somewhat more difficult. However, in Ref. [1] the solution is given in terms of the error function erf and the exponential integral Ei:

$$\int_0^\infty \mathrm{d}r'\, \frac{e^{-cr'^2}}{r'+x} = e^{-cx^2}\left[\frac{\pi}{2\imath}\mathrm{erf}(\imath\sqrt{cx^2}) - \frac{1}{2}\mathrm{Ei}(cx^2)\right] \,. \tag{3.10}$$

As $c \sim 1/Dt$ in the considered case, erf and Ei can be expanded for $c \ll 1$ at reasonably large t:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)} \qquad \Rightarrow \qquad \frac{\mathrm{erf}(\imath z)}{\imath} = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{z^{2n+1}}{n!(2n+1)} \,, \tag{3.11}$$

$$\mathrm{Ei}(z) = \gamma + \ln z + \sum_{n=1}^{\infty} \frac{z^n}{n\, n!} \qquad \text{for } z > 0 \,, \tag{3.12}$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. Applying the previous findings to integrate Eq. 3.8 with $c = 1/2Dt$ and $x = r_0$ we finally get:

$$\begin{aligned} f(t) &= -\int_{r_0}^{\infty} \mathrm{d}r'\, \dot{p}(r',t) \\ &\approx -\frac{r_0\varepsilon}{\sqrt{8\pi D^3}} \frac{1}{\sqrt{t^5}} \Bigg\{ -2r_0 + \sqrt{\frac{2\pi}{Dt}}(r_0^2 - 2Dt) + \left(6r_0 - \frac{2r_0^3}{Dt}\right) \exp\left[-\chi^2\right] \times \\ &\qquad \left[\sqrt{\pi} \sum_{n=0}^{\infty} \frac{1}{n!(2n+1)} \chi^{2n+1} - \frac{1}{2}\left(\gamma + 2\ln\chi + \sum_{n=1}^{\infty} \frac{1}{n\, n!} \chi^{2n}\right)\right] \Bigg\} \,, \end{aligned} \tag{3.13}$$

with $\chi = r_0/\sqrt{2Dt}$. A further expansion in terms of t is difficult, since $\ln\chi \sim \ln(1/t)$ diverges for both large and small t. However, since this logarithm is not contained in the leading contribution, the asymptotic behavior of the distribution of return times can still be calculated:

$$f(t) \approx -\frac{r_0\varepsilon}{\sqrt{8\pi D^3}}\frac{1}{\sqrt{t^5}}\sqrt{\frac{2\pi}{Dt}}(-2Dt) = \frac{r_0\varepsilon}{Dt^2}\ . \tag{3.14}$$

This asymptotic behavior is somewhat surprising recalling the work of Pólya [112] which predicts a power law of $f(t) \sim t^{-3/2}$ instead of $f(t) \sim t^{-2}$. Fig. 3.9a shows the result of two different numeric solutions for parameters $a = r_0 - \varepsilon$ and D according to barnase and barstar with $\varepsilon = 0.1\text{Å}$, as well as their relative deviation and data from the corresponding simulation for comparison (see Fig. 3.6b). The first solution was obtained via the numerical integration of Eq. (3.7). The second is the approximate solution Eq. (3.13) with a reasonably high number of summands. The plot also shows the relative deviation of these two types of solutions $\mathrm{abs}(f_1(t) - f_2(t)/f_1(t))$, which is well below 1% over the whole range of the plot. This verifies that the assumptions we made to find the approximate solution Eq. (3.13) are reasonable. However, as already discussed, the crossover to a t^{-2} behavior is not seen in the simulation data. On the other hand, the histogram obtained by simulation shows an exponential decay at some point, which is due to the finite available space for exploration due to the periodic boundary conditions, which we are not able to consider in the analytic approach.

Resting time at translational encounter

In Ref. [34] an even more general solution for heat equations with spherical symmetry is given, which involves two generally parameterizable radiation boundary conditions. As explained in the beginning of this section, the problem of resting time T_{on} involves a reflecting boundary at $b = R_1 + R_2$ and an absorbing boundary at $a = b + r_1^p + r_2^p$. In the following we again write t instead of T_{on} for the situation considered in the analytic solution. For these particular conditions and applying an initial "heat" distribution $f(r') = \delta(r' - r_0)$, the solution of the probability density $p(r,t)$ as introduced above is given by:

$$p(r,t) = \frac{2}{r}\sum_{n=1}^{\infty} e^{-D\alpha_n^2 t/2} r_0 R_n(r) R_n(r_0)\ , \tag{3.15}$$

$$\text{with} \qquad R_n(r) = \frac{\sin\left((r-a)\alpha_n\right) + a\alpha_n\cos\left((r-a)\alpha_n\right)}{\sqrt{a^2\alpha_n^2(b-a)+b}}\ . \tag{3.16}$$

Particularly, α_n are the solutions of the equation $a\alpha = \tan\left((a-b)\alpha\right)$. The negative time derivative of the probability distribution is thus:

$$-\dot{p}(r,t) = \sum_{n=1}^{\infty} A_n e^{-D\alpha_n^2 t/2}\frac{1}{r}\left\{\sin\left((r-a)\alpha_n\right) + a\alpha_n\cos\left((r-a)\alpha_n\right)\right\}\ , \tag{3.17}$$

$$\text{with} \qquad A_n = \frac{r_0 R_n(r_0) D\alpha_n^2}{\sqrt{a^2\alpha_n^2(b-a)+b}}\ . \tag{3.18}$$

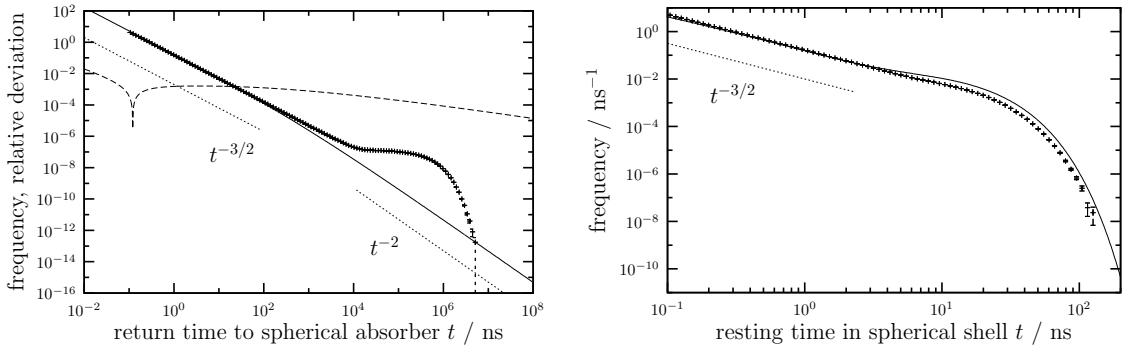


Figure 3.9: (a) Numerical solution of the return time problem according to Eq. (3.7) and Eq. (3.13) (solid line) and their relative deviation (dashed line) as well as indicated scaling behavior (dotted lines) and simulation data (data points). A general normalization of the numerical solution is not available due to the divergence at $t \to 0$ and is thus obtained by a fit to the simulation data. (b) Numerical solution of the resting time problem according to Eq. (3.17) (solid line) with the scaling behavior indicated by the dotted line and compared to the according simulation data (data points).

Unfortunately, the integration $f(t) = -\int_a^b \mathrm{d}r'\, \dot{p}(r', t)$ is a complicated task as the r dependent $\sin$- and $\cos$-terms can be heavily fluctuating due to possibly large α_n. Thus it can only be evaluated numerically. However, the structure of the solution is again an infinite sum of exponentials with respect to time, where the mentioned integration with respect to r acts on the third factor in the writing of Eq. (3.17) and leads to combined coefficients $A_n \times \int$. Fig. 3.9b shows the comparison of this analytic solution parameterized with a, b and D according to the barnase and barstar system and the data obtained by simulations as already shown in Fig. 3.6b. The initial radius r_0 was chosen to be $a - 0.1\text{Å}$ and the normalization of the numerical solution was again fitted to the simulation data. The plot shows a good agreement and especially as here the finiteness of the system is directly incorporated in the analytic approach the exponential decay of the simulation data is well reproduced.

3.4 Discussion

The main goal of this chapter is to model protein encounter in a generic framework which later will allow us to study the dynamics of protein clusters. Our model approach incorporates structural, electrostatic and hydrodynamic properties of the considered proteins. These are thought to be the major factors governing protein encounter. Not included are conformational changes of the proteins after association, related entropic terms, and the molecular nature of the surrounding solvent that becomes relevant at close distances [2]. The model parameters are extracted from the atomic structures available in the PDB by generally applicable protocols as described in chapter 2 and Sect. 3.1. In principle, these methods of data extraction can be fully automatized and are thus capable of producing well

comparable parameter sets.

The biggest advantage of this coarse-grained model is the possibility to extend the simulations to large scales in terms of particle numbers, time and system size. In many of the earlier studies [51, 100, 101, 172] the system was prepared already close to encounter and the overall association rate was then calculated via a sophisticated path-integral like procedure. In contrast, our simulations account for the whole process of diffusional encounter and is thus rather general, allowing for spanning large time scales via our adaptive time step algorithm. In particular, each set of simulations consists of 10^4 to 10^5 runs of lengths up to the order of seconds and was performed on a standard CPU within hours of computer time.

Being able to directly obtain the first passage times (FPT) of the encounter processes in our model allows to check the validity of several phenomenological assumptions. First of all, the FPT distribution matched very well a Poisson process with a single stochastic rate, as seen in Fig. 3.2a, which validates the notions of encounter and association. Moreover, both ways of controlling the particle density, particularly by changing the boundary box size and by varying the number of particles in the box, are shown to match the expected scaling behavior. Therefore we can conclude that the computational model studied here fulfills the general requirement of stochastic bimolecular association processes that describe binding by a single rate constant. Furthermore, three well-known bimolecular systems with different characteristics were purposely chosen to check whether our effective representations of the protein properties would lead to reasonable and significantly distinguishable results. Indeed, this is the case as the discussion of the results in Tab. 3.2 in the respective section shows.

When comparing the results for the encounter rates in Tab. 3.2 with previous studies from the field of bimolecular protein association, several aspects have to be kept in mind. First, throughout this study, we do only consider the *encounter* of our model particles. The complete association of the complex still lacks the step over a final free energy barrier, which is due to effects such as the dehydration of the protein surfaces, completion of the alignment involving a loss of entropy and many other processes which have been studied in detail in Ref. [2]. In the framework of this study, this final step could be modeled by a stochastic rate criterion, where the rate can be obtained by transition state theory from the energy landscapes characterized in atomistic calculations. In any case, any additional process to be included can only lower the rates found in our study. Thus, our results can be considered as an upper boundary.

In the work on barnase:barstar by Schreiber et al. [126], the authors reported that the association between barnase and barstar is a diffusion-limited reaction. The argument for this is that the association rates at high ionic concentrations in the solution, i.e. for the limit in which the electrostatic steering gets negligible, are clearly lowered by the addition of glycerol, which will lead to slower diffusion. Assuming diffusion control, the reactive step over the final barrier should be kinetically unimportant, as generally discussed in Sect. 1.3. Indeed, we see that our results for the encounter rates lead to values in the correct order of magnitude of $k \approx 10^9 \mathrm{M}^{-1}\mathrm{s}^{-1}$, which is similar to the experimental value obtained by Schreiber et al. for the association constant of barnase:barstar $k = 8 \cdot 10^8 \mathrm{M}^{-1}\mathrm{s}^{-1}$ and more

[125]. However, the basal association rate, i.e., the rate at high ionic strength in the limit of low Coulomb interaction is given to be about $k_{\mathrm{basal}} < 10^6\mathrm{M}^{-1}\mathrm{s}^{-1}$ in experiments [126]. Given that the association process of brn:brs is diffusion limited, these findings should actually agree with our values for $\mathcal{M}1$. But as we already discussed the influence of the effective electrostatics introduced in $\mathcal{M}2$ do not result in such a drastic change of the encounter behavior.

In several earlier approaches, similar problems have been addressed by computational and analytical studies. In different works by Zhou, basal encounter rates for particles with reactive patches have been found to be $k_{\mathrm{basal}} = 4{\cdot}10^6\mathrm{M}^{-1}\mathrm{s}^{-1}$ [170] and $k_{\mathrm{basal}} = 10^7\mathrm{M}^{-1}\mathrm{s}^{-1}$ [171]. Note that in both cases the patches were flat areas above the surface of the spherical model particles, which had a smaller angular extension compared to our cases, and especially required a much closer translational approach (0.7Å in [170]) to form the encounter. If we expand the graph in Fig. 3.4a to smaller patch radii like $r = 1\text{Å}$, we also find basal rates in the order of $k = 10^7\mathrm{M}^{-1}\mathrm{s}^{-1}$. Also, the deviation between $\mathcal{M}1$ and $\mathcal{M}2$, i.e. the impact of the effective electrostatics is more prominent and could enhance the encounter rate by about two orders of magnitude, which is consistent with the findings in the previously cited work. There, the effect of Coulombic interaction is reflected with a Boltzmann-like factor due to a pairwise Coulomb energy. This approach works well, as shown in Ref. [172], and has been recently used in a more complex model study of the energy landscape of protein-protein association [11, 12].

However, the aim of this work is not to model the biomolecular details of the association process, but to find a suitable link of a coarse-grained model for the process of diffusional encounter to a more sophisticated treatment of the final complexation as it has been described by Helms and co-workers [3, 140]. This is reasonable, as many effects become important for very close approach of proteins. Furthermore, the hydration shell has a typical thickness of about 3Å and will therefore in principle hinder the approach of two proteins to distances below 6Å. Moreover, all of the considered protein systems feature distinct key-lock binding interfaces regarding the steric structure, apart from some flexibility due to intrinsic thermal motion. Therefore, it makes sense to represent the encounter area by a three-dimensionally spread object rather than a flat surface region. Indeed, the results of our studies show that a generic approach like ours is capable of reproducing encounter rates in a reasonable order of magnitude, qualitatively replicating general expected features. Our investigation of the dependency of the encounter rate and the patch radius shows that the choice of the geometry of the reactive area is at least as crucial for the results as the definition of the model interactions and its parameters. In principle, one can think of the patch radius as a valuable tuning parameter to fit experimental results and the encounter kinetics in the computational model. The observation furthermore approves the importance of rotational diffusion as an accelerating mechanism with respect to the aligning of the proteins. We show that the encounter rate decreases not as slow for smaller patch sizes as static geometric arguments would suggest.

Our approach makes it possible to investigate general features of the encounter process. In particular, we dissect the pathway to the encounter complex in several levels of alignment between our model proteins. As we observe the

full trajectory to encounter in our simulations, we are able to extract the number of close approaches (contacts) N between the proteins until they finally reach a reasonably aligned state to bind. The distribution of N is again in all cases well described by a single exponential decay. This behavior is not obvious as the probability of success for one contact is depending on several aspects of diffusion in a complex manner. First, the closer the rotational alignment is to the encounter state at the beginning of the contact, the higher is the probability of success. Second, this initial alignment is also coupled to the last contact if the time in between, T_{off}, is small. Finally, longer contact resting times T_{on} also increase the probability of encounter. It is interesting, that all these effects still lead to a simple Poisson distribution of the number of contacts N when averaging over the initial conditions as it is done in this study. Furthermore, we find that the distributions of these resting and return times cannot be described by a Poisson process, but are consistent with the expectations for a spatially constricted random walk in three dimensions. We find that the particular mean FPT to encounter scales linearly to the number of unsuccessful contacts in most cases. This seems to be a very fundamental qualitative feature irrespective of the details of the proteins and the applied model. However, for cytc:ccp the behavior is qualitatively different, which is consistent with earlier studies of this highly electrostatically steered complex. There, the encounter rate k is virtually independent of $\langle N \rangle$ and the choice of the patch radius r if the Coulombic attraction is effectively modeled, i.e. in our approaches $\mathcal{M}2$ and $\mathcal{M}3$. The distributions of resting and return times can be motivated by analytic approximations. Especially for the resting times T_{on} the analytic results show good agreement with the simulation data.

Here, we have shown that the relevant features of bimolecular protein association on length and time scales above the atomistic level, i.e., in the range of nanometers and nanoseconds, are reproduced well. Due to its computational efficiency our model is an excellent starting point for studying the dynamics of larger protein complexes, if importance is attached to aspects of association pathways which happen on scales of milliseconds to seconds. Unlike in more detailed simulation models, we are able to gather significant statistics also over complete trajectories on these timescales.

An interesting subject for future work is the direct evaluation of the entropic contribution to the free energy in our model. Particularly, this would involve the analysis of distance and position histograms over long simulation trajectories as it has been done by Spaar et al. [141] in a particular distance regime. With our generic model it is possible to compute similar entropy landscapes for larger distances and especially to take the effect of finite concentration into account, which was not possible in the cited study.

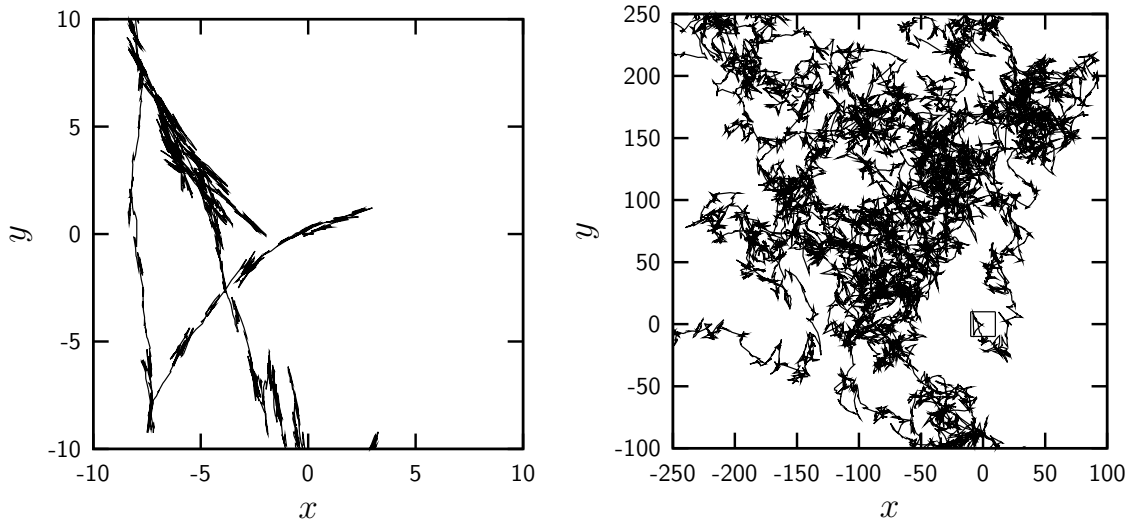
Chapter 4

Influence of hydrodynamic anisotropy

In the previous chapter we investigated the effects of electrostatic and steric interactions on rates of molecular encounter for spherical or close-to-spherical particles. Although many proteins are globular, some are not (compare Fig. 3.1). Even more importantly, clusters of globular proteins are not globular. As the encounter process basically denotes the transport part of protein interaction, it depends on the particular diffusion properties of the interacting objects and thus sufficiently strong anisotropy might affect the encounter rate. In this chapter we want to investigate this aspect in more detail. First, the features distinguishing anisotropic from isotropic diffusion are quantified. Particularly, we calculate the relevant time scales on which diffusion can be anisotropic. Then, the encounter dynamics of ellipsoidal molecular particles is observed in detail. The influence of the aspect ratio of the ellipsoids on the encounter rates is evaluated in computer simulations similar to those in chapter 3. By investigating the altered accessibility of the encounter patches due to steric constraints we find that this effect does mainly account for the observed variation in encounter rates. Finally, we compare simulations with and without consideration of the anisotropic diffusivity to gain a better understanding for its importance.

4.1 From anisotropic to isotropic diffusion

Objects exhibiting anisotropic diffusion properties show qualitatively different trajectories on a small time scale compared to isotropically moving objects. An anisotropic random walk started at some particular orientation at time t_0 will initially show a preferred direction of motion. As the particle will also undergo rotational diffusion in general, this principal direction will change over time. A typical trajectory is shown in Fig. 4.1a. The memory of the initial orientation will get lost due to the stochastic rotations after some time. That is, for large times there cannot be a preferred direction anymore and each direction of motion is equally probable. Therefore, diffusional anisotropy can only exist up to a certain scale and there has to be a crossover to isotropic diffusion on large scales as illustrated



(a) Small time and length scale **(b)** Large time and length scale

Figure 4.1: Dimensionless example of two-dimensional trajectories of a particle performing a random walk with principal diffusion coefficients $D^t_{xx} = 10$ and $D^t_{yy} = 0.1$ for translation and $D^r = 0.1$ for rotation at different resolution: (a) Closer look at the starting point of the trajectory $x_0 = y_0 = 0$ with a time resolution $\Delta t = 10^{-2}$; (b) Coarser time resolution of $\Delta t = 1$ and a larger viewing window. The area shown in (a) is indicated by a dashed square.

in Fig. 4.1b. In Ref. [75] Han and colleagues experimentally observed the two-dimensional diffusion of an ellipsoidal particle in a thin glass cell. They found nice agreement between analytic calculations and corresponding experiments. The calculation of Han et al. deals with a two-dimensional system where the state of a particle is fully described by two translational (x, y) and one rotational degree of freedom (θ). No diffusional coupling is considered, i.e., the diffusion matrix does not contain off-diagonal entries:

$$\mathbb{D} = \begin{pmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_\theta \end{pmatrix} . \tag{4.1}$$

In $\mathbb{D}$ the first two components denote the two translational degrees of freedom in the plane while the third component denotes rotations in the plane. The action of a rotation by angle θ is written in a matrix form:

$$\mathbb{S} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{4.2}$$

The transformed diffusion matrix $\bar{\mathbb{D}}$ reads:

$$\bar{\mathbb{D}} = \mathbb{S}\mathbb{D}\mathbb{S}^{\mathsf{T}} = \begin{pmatrix} D_+ & 0 & 0 \\ 0 & D_+ & 0 \\ 0 & 0 & D_\theta \end{pmatrix} + D_- \begin{pmatrix} \cos(2\theta) & \sin(2\theta) & 0 \\ \sin(2\theta) & -\cos(2\theta) & 0 \\ 0 & 0 & 0 \end{pmatrix} , \tag{4.3}$$

with $D_+ = (D_x + D_y)/2$ and $D_- = (D_x - D_y)/2$. For a Langevin process without drift in the over-damped limit it is (compare chapter 2):

$$\frac{\partial}{\partial t}\mathbf{x}_t = \mathbf{g}_t , \qquad \langle \mathbf{g}_t \rangle = 0 , \qquad \langle \mathbf{g}_t \mathbf{g}_{t'} \rangle = 2\mathbb{D}\delta(t - t') . \tag{4.4}$$

If the object has orientation θ at time t, then for a small time interval δt the mean square displacement is:

$$\langle x(t+\delta t)^2 \rangle = \langle x(t)^2 \rangle + 2\langle (D_+ + D_- \cos(2\theta)) \rangle_t \delta t , \tag{4.5}$$
$$\langle y(t+\delta t)^2 \rangle = \langle y(t)^2 \rangle + 2\langle (D_+ - D_- \cos(2\theta)) \rangle_t \delta t . \tag{4.6}$$

The average in the second summands considers all possible orientations weighted by their probability at time t due to rotational diffusion. Assuming $\theta_0 = t_0 = 0$ this leads to:

$$\langle (D_+ \pm D_- \cos(2\theta)) \rangle_t = D_+ \pm D_- \int_{-\infty}^{\infty} \mathrm{d}\theta \, p(\theta, t) \cos(2\theta) = D_+ \pm D_- e^{-4D_\theta t} , \tag{4.7}$$

$$p(\theta, t) = \frac{1}{\sqrt{4\pi D_\theta t}} e^{-\theta^2/(4D_\theta t)} . \tag{4.8}$$

Hence, the total mean square displacement at time t is:

$$\langle x(t)^2 \rangle = 2 \int_0^t \mathrm{d}t' \, \langle (D_+ + D_- \cos(2\theta)) \rangle_{t'} = 2D_+ t + \frac{D_-}{2D_\theta} \left(1 - e^{-4D_\theta t}\right) . \tag{4.9}$$

We conclude that the characteristic time scale for the crossover from anisotropic to isotropic diffusion in two dimensions is $1/4D_\theta$. In case of diffusional coupling between translational and rotational degrees of freedom the above calculations are not applicable due to a more complex structure of the diffusion matrix after rotation.

4.1.1 Crossover in three dimensions

In three dimensions the situation is more complicated as there are now three different rotation axis, which we will consider as independent (no coupling). This

makes the diffusion matrix a 6×6 matrix:

$$\mathbb{D} = \begin{pmatrix} D_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & D_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & D_1^\theta & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^\theta & 0 \\ 0 & 0 & 0 & 0 & 0 & D_3^\theta \end{pmatrix} . \tag{4.10}$$

The torque due to the rotational diffusion basically leads to a rotation of the system around a vector whose component along the unit vector i (in the body fixed coordinate system) is drawn from a Gaussian distribution with the width D_i^θ, while the length of the rotation vector determines the magnitude of the rotation angle. As the rotations are supposed to be completely independent, we only have to consider the upper left 3×3 part of the diffusion matrix, which is responsible for the translational diffusion:

$$\mathbb{D}^t = \begin{pmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{pmatrix} . \tag{4.11}$$

A rotation of a 3×3 matrix determined by a vector of angles $\theta = (\theta_1, \theta_2, \theta_3)^\dagger$ around the three principal axes can be described by the following rotation matrix:

$$\mathbb{S} = e^{-\theta \cdot \mathbf{J}} , \qquad \mathbf{J} = \left(\mathbb{J}^1, \mathbb{J}^2, \mathbb{J}^3\right) . \tag{4.12}$$

$\theta \cdot \mathbf{J}$ denotes a formal scalar product and $\mathbb{J}^k$ are matrices defined by $\mathbb{J}^k_{ij} = \epsilon_{ikj}$, where ϵ is the Levi-Civita symbol. We proceed considering only small rotations occurring at small times, so that we can expand the rotation in orders of θ:

$$\mathbb{S}_{ij} = \delta_{ij} \left(1 - \frac{1}{2}\sum_{k=1}^{3} \theta_k^2\right) + \epsilon_{ikj}\theta_k \left(1 - \frac{1}{6}\sum_{l=1}^{3} \theta_l^2\right) + \frac{1}{2}\left(\theta_i\theta_j\right) + \mathcal{O}(\theta^4) . \tag{4.13}$$

By applying the rotation to $\mathbb{D}^t$ we get $\bar{\mathbb{D}^t} = \mathbb{S}\mathbb{D}^t\mathbb{S}^\mathsf{T}$, considering only terms up to second order in θ:

$$\begin{aligned} \bar{\mathbb{D}^t}_{ij} = \delta_{ij} &\left[D_i + \left(\sum_{k=1}^{3} \theta_k^2\right)\left(-2D_i + \sum_{m=1}^{3} D_m\right) - \sum_{m=1}^{3} \theta_m^2 D_m\right] \\ &+ \theta_i\theta_j \left(\frac{3}{2}(D_i + D_j) - \sum_{m=1}^{3} D_m\right) + \epsilon_{ikj}\theta_k \left(D_j - D_i\right) + \mathcal{O}(\theta^3) . \end{aligned} \tag{4.14}$$

As in the first section, the non-diagonal entries are odd in θ_i and thus do not give a contribution when averaging over the possible orientations, weighted by their

probability. Again, the average of the diagonal entries is the central quantity for the calculation of the mean square displacement. As we already did an expansion in θ and we also assumed $t \ll 1$, it will now also be sufficient to consider a probability distribution of angles θ_i due to rotational diffusion, which is correct up to second order in t. Therefore, the Gaussian probability distribution from the first section can be replaced by a uniform distribution regarding correct integral boundaries. The particularly interesting average $\langle \bar{\mathbb{D}}^t{}_{ii} \rangle_t$ is then:

$$\langle \bar{\mathbb{D}}^t{}_{ii} \rangle_t = \frac{1}{w_1 w_2 w_3} \int_{-w_1/2}^{w_1/2} \mathrm{d}\theta_1 \int_{-w_2/2}^{w_2/2} \mathrm{d}\theta_2 \int_{-w_3/2}^{w_3/2} \mathrm{d}\theta_3 \bar{\mathbb{D}}^t{}_{ii} \,, \tag{4.15}$$

where $w_i = \sqrt{24 D_i^\theta t}$ is the width of the uniform distribution interval. As already stated, the terms odd in a particular θ_i will not give a contribution. That is, the integral in Eq. 4.15 will only lead to zeroth and second moments of the angular distribution. Hence, its solution up to linear order in t is:

$$\langle \bar{\mathbb{D}}^t{}_{ii} \rangle_t \approx D_i \left[1 - 4 \left(\sum_{k=1}^{3} D_k^\theta t \right) + 6 D_i^\theta t \right] + \sum_{m=1}^{3} 2 D_m t \left[\left(\sum_{k=1}^{3} D_k^\theta \right) - D_i^\theta - D_m^\theta \right] \,. \tag{4.16}$$

Therefore, by the average action of the rotational diffusion, D_i transforms into an effective translational diffusion constant $\langle \bar{\mathbb{D}}^t{}_{ii} \rangle_t$ over time t. Considering a vector with the three principal diffusion coefficients $\mathbf{D} = (D_1, D_2, D_3)$, the evolution of the effective, orientation averaged vector of diffusion coefficients $\langle \mathbf{D} \rangle(t) = (\langle \bar{\mathbb{D}}^t{}_{11} \rangle_t, \langle \bar{\mathbb{D}}^t{}_{22} \rangle_t, \langle \bar{\mathbb{D}}^t{}_{33} \rangle_t)$ can be expressed in a matrix form, not taking into account non-linear terms in t:

$$\langle \mathbf{D} \rangle(t) = \mathbf{D} \cdot \mathbb{R}(t) + \mathcal{O}(t^2) \,, \tag{4.17}$$

$$\mathbb{R}(t) = \mathbb{I} + 2t \begin{pmatrix} -(D_2^\theta + D_3^\theta) & D_3^\theta & D_2^\theta \\ D_3^\theta & -(D_1^\theta + D_3^\theta) & D_1^\theta \\ D_2^\theta & D_1^\theta & -(D_1^\theta + D_2^\theta) \end{pmatrix} \,. \tag{4.18}$$

The principal axes of effective motion are constant since the coupling terms between the translational degrees of freedom vanish when averaging over all possible orientations. Because rotational diffusion is independent of time and orientation, which makes $\mathbb{R}$ a constant property of the diffusion of some arbitrary body, we can also apply $\mathbb{R}(\delta t)$ to some diffusion vector $\langle \mathbf{D} \rangle(t)$ and get:

$$\langle \mathbf{D} \rangle(t + \delta t) = \langle \mathbf{D} \rangle(t) \cdot \mathbb{R}(\delta t) + \mathcal{O}(\delta t^2) \,. \tag{4.19}$$

Thus, it is possible to evaluate the effective change of the diffusion coefficients for large times t in small steps δt with only making errors of $\mathcal{O}(\delta t^2)$:

$$\langle \mathbf{D} \rangle(N \delta t) = \mathbf{D} \cdot (\mathbb{R}(\delta t))^N + \mathcal{O}(N \delta t^2) \,. \tag{4.20}$$

The potentiation of $\mathbb{R}$ is easiest when using a diagonal form:

$$\mathbb{R}(\delta t) = \mathbb{S}(\delta t) \cdot \mathbb{R}_d(\delta t) \cdot \mathbb{S}(\delta t)^{-1} \,, \tag{4.21}$$

with:

$$\mathbb{R}_d(\delta t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-2\delta t(c+d) & 0 \\ 0 & 0 & 1-2\delta t(c-d) \end{pmatrix} , \tag{4.22}$$

$$\mathbb{S}(\delta t) = \mathbb{S} = \begin{pmatrix} c & a+d & a-d \\ c & -c-a-d & -c-a+d \\ c & c & c \end{pmatrix} , \tag{4.23}$$

$$a = D_3^\theta - D_1^\theta , \qquad b = D_1^\theta - D_2^\theta , \qquad c = D_1^\theta + D_2^\theta + D_3^\theta , \tag{4.24}$$

$$d = \sqrt{D_1^{\theta 2} + D_2^{\theta 2} + D_3^{\theta 2} - D_1^\theta D_2^\theta - D_1^\theta D_3^\theta - D_2^\theta D_3^\theta} . \tag{4.25}$$

As the transformation into a diagonal form is not unique, an asymmetry is introduced at this step with respect to the different components of $\mathbb{D}$. One would actually expect that the three terms a, b and c should have a similar form. However, in our choice c has a different structure than a and b. We now go to the limit of large N with $\delta t = t/N$. This limit is meaningful as the error in Eq. (4.20) vanishes: $\mathcal{O}(Nt^2/N^2) = \mathcal{O}(t^2/N) \to 0$. This basically means that we ask for $\langle \mathbf{D} \rangle(t)$ with infinite accuracy. As $\mathbb{S}$ is independent of δt, the limit does only apply to $(\mathbb{R}_d(\delta t))^N$, where we find:

$$\lim_{N\to\infty} \left(1 - 2\frac{t}{N}(c \pm d)\right)^N = e^{-2(c\pm d)t} . \tag{4.26}$$

This leads to an exact expression for $\langle \mathbf{D} \rangle(t)$:

$$\langle \mathbf{D} \rangle(t) = \mathbf{D} \cdot \mathbb{S} \cdot \mathbb{R}_d^\infty(t) \cdot \mathbb{S}^{-1} , \quad \text{with} \quad \mathbb{R}_d^\infty(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-2(c+d)t} & 0 \\ 0 & 0 & e^{-2(c-d)t} \end{pmatrix} , \tag{4.27}$$

which can be evaluated to:

$$\langle \mathbf{D} \rangle(t) = \begin{pmatrix} f + (ah_2 - bh_3)\tilde{g}_- + dh_1\tilde{g}_+ \\ f + (ah_1 - bh_2)\tilde{g}_- + dh_2\tilde{g}_+ \\ f + (ah_3 - bh_1)\tilde{g}_- + dh_3\tilde{g}_+ \end{pmatrix} , \tag{4.28}$$

$$\text{with} \quad f = \frac{D_1 + D_2 + D_3}{3} , \quad \tilde{g}_\pm = \frac{e^{-2(c-d)t} \pm e^{-2(c+d)t}}{6d} , \quad h_i = 3(D_i - f) . \tag{4.29}$$

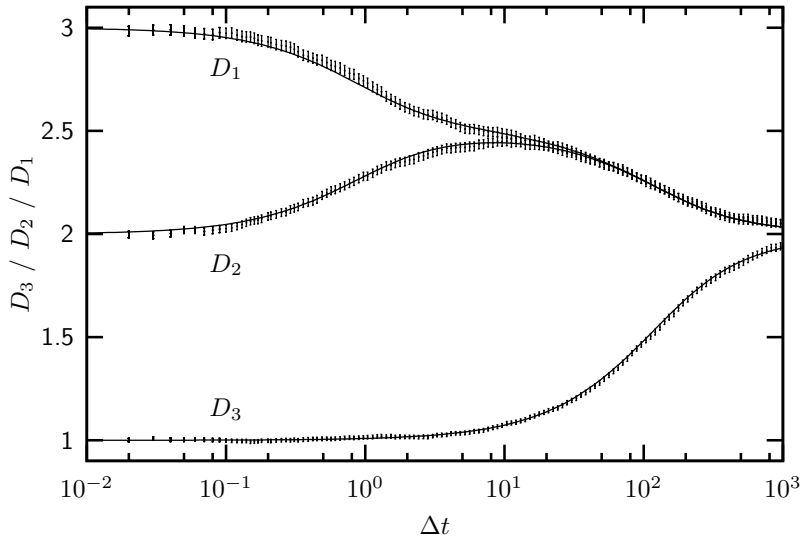


Figure 4.2: Example for the evolution of the effective principal translational diffusion coefficients for the following choice of parameters: $D_1 = 3$, $D_2 = 2$, $D_3 = 1$, $D_1^\theta = 0$, $D_2^\theta = 0.005$, $D_3^\theta = 0.5$. The data points have been obtained by simulation (the error bars depict the standard deviation obtained from 10^5 individual runs), the line represents the theoretic prediction.

In analogy to the two-dimensional case, we can now evaluate the overall, orientation averaged mean square displacement by integrating over $\langle \mathbf{D} \rangle(t)$:

$$\langle \mathbf{x}(t)^2 \rangle = 2 \int_0^t \mathrm{d}t' \, \langle \mathbf{D} \rangle(t') = \begin{pmatrix} 2ft + (ah_2 - bh_3)g_- + dh_1 g_+ \\ 2ft + (ah_1 - bh_2)g_- + dh_2 g_+ \\ 2ft + (ah_3 - bh_1)g_- + dh_3 g_+ \end{pmatrix} , \tag{4.30}$$

$$\text{with} \qquad g_\pm = 2 \int_0^t \mathrm{d}t' \, \tilde{g}_\pm = \frac{1 - e^{-2(c-d)t}}{6(c-d)d} \pm \frac{1 - e^{-2(c+d)t}}{6(c+d)d} . \tag{4.31}$$

This can also be used to obtain effective translational diffusion coefficients via $\mathbf{D}^{\mathrm{eff}}(t) = \langle \mathbf{x}^2(t) \rangle / 2t$. Fig. 4.2 shows the effective principal diffusion coefficients for an arbitrary object which cannot rotate around the x_1-axis. This choice is unnatural but leads to a clearer understanding of the plot. As rotational diffusion is much faster around the x_3-axis than around the x_2-axis, the anisotropy in the x_1-x_2-plane is lost first due to D_3^θ. For larger Δt this average $D_{1,2} = (D_1 + D_2)/2$ is also mixed with D_3 due to the slower rotations according to D_2^θ. The analytic solution matches the simulation results very well. As in the two-dimensional case, the additional consideration of diffusional coupling does not lead to a closed solution.

The effective diffusion coefficients $\mathbf{D}^{\mathrm{eff}}(t)$ should be equal to the initial ones $\mathbf{D}$ for $t \to 0$. The limit of $g_\pm/(2t)$ for $t \to 0$ can be determined by expanding the

exponential functions up to first order:

$$\lim_{t\to 0}\frac{g_\pm}{2t} = \frac{2(c-d)t}{12(c-d)dt} \pm \frac{2(c+d)t}{12(c+d)dt} = \begin{cases} 0 & \text{for } g_- \\ \dfrac{1}{3d} & \text{for } g_+ \end{cases} . \tag{4.32}$$

Therefore, $D_i^{\text{eff}}(t)$ behaves as expected for small t:

$$\lim_{t\to 0} D_i^{\text{eff}}(t) = f + \frac{1}{3d}dh_i = D_i \, . \tag{4.33}$$

In contrast, we expect the crossover to isotropic diffusion for $t \to \infty$, i.e., all $D_i^{\text{eff}}(t)$ should be equal in this limit. Indeed, this is the case as the prefactors of $g_\pm$ do not depend on t and $g_\pm$ approach finite values for large t:

$$\mathbf{D}^{\text{eff}}(t) = \frac{1}{2t}\langle \mathbf{x}(t)^2\rangle = f + \frac{C_i g_- + dh_i g_+}{2t} \overset{t\to\infty}{\longrightarrow} f \, , \tag{4.34}$$

where C_i just denotes the prefactor of g_- according to Eq. (4.30). Therefore, the analytic limits of the main result agree with our expectations for $\mathbf{D}^{\text{eff}}(t)$.

4.2 Brownian motion of rigid ellipsoidal particles

Diffusion properties

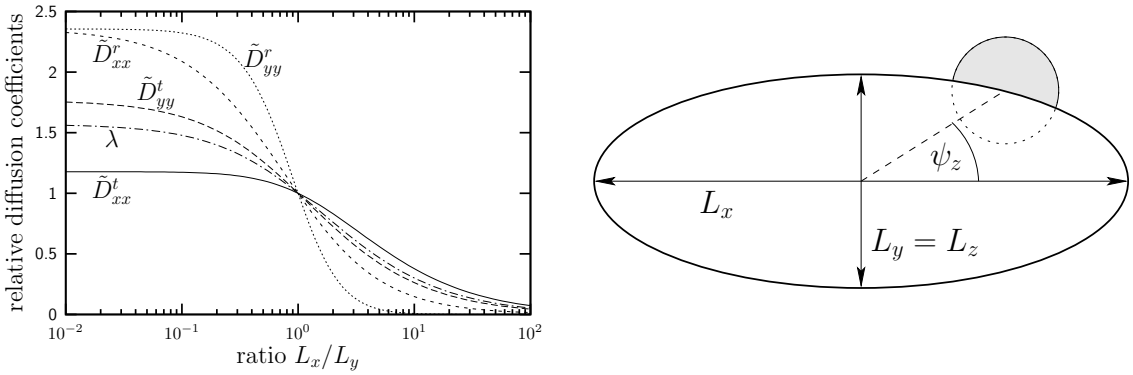
In our study we want to observe the encounter of anisotropic particles. Using ellipsoids with two equal semi axes is advantageous for two reasons. First, this type of ellipsoid is one of the few geometric bodies for which closed analytic expressions of the friction coefficients are known, which have been calculated by Perrin [110, 111]. The results are summarized in Ref. [134]. Assuming the ellipsoid is viewed in a coordinate system spanned by the principal axes, it is symmetric with respect to mirroring in all normal planes and, therefore, the friction matrix has only diagonal entries. Letting the x-axis be the rotation axis, i.e. the diameter of the spheroid in y and z direction are equal $L_y = L_z$, these entries are:

$$\zeta^t_{xx} = 8\pi\eta\frac{L_x^2 - L_y^2}{(2L_x^2 - L_y^2)S - 2L_x} \, , \tag{4.35}$$

$$\zeta^t_{yy} = \zeta^t_{zz} = 16\pi\eta\frac{L_x^2 - L_y^2}{(2L_x^2 - 3L_y^2)S + 2L_x} \, , \tag{4.36}$$

$$\zeta^r_{xx} = \frac{4\pi}{3}\eta\frac{(L_x^2 - L_y^2)L_y^2}{2L_x - L_y^2 S} \, , \tag{4.37}$$

$$\zeta^r_{yy} = \zeta^r_{zz} = \frac{4\pi}{3}\eta\frac{L_x^4 - L_y^4}{(2L_x^2 - L_y^2)S - 2L_x} \, . \tag{4.38}$$



(a) **(b)**

Figure 4.3: (a) Relative change of the diffusion coefficients and the relative mobility measure λ (chain dotted line) for an ellipsoid with varying diameter along the symmetry axis normalized by the corresponding values for a sphere. (b) Basic geometry of the considered ellipsoid with a reactive patch on its surface at some angle ψ_z with respect to the symmetry axis $\mathbf{e}_x$.

Here, η is the viscosity and S is a function of L_x and L_y which is different for prolates and oblates:

$$S = \begin{cases} \dfrac{2}{\sqrt{L_x^2 - L_y^2}} \log \dfrac{L_x + \sqrt{L_x^2 - L_y^2}}{L_y} & \text{, if } L_x \geq L_y \text{ ,} \\ \dfrac{2}{\sqrt{L_y^2 - L_x^2}} \arctan \dfrac{\sqrt{L_y^2 - L_x^2}}{L_x} & \text{, if } L_x < L_y \text{ .} \end{cases} \tag{4.39}$$

Note that in the original reference the expressions are given in terms of the semi axes, which results in different prefactors. The friction coefficients for general ellipsoids with three distinct semi axes involve elliptic integrals and can thus only be evaluated numerically. We furthermore introduce a relative mobility measure λ which compares the trace of the translational part of the mobility matrix according to the inverse of Eq. (4.35) and Eq. (4.36) with the spherical case $\zeta^t_{xx} = \zeta^t_{yy} = 3\pi\eta L_y$:

$$\lambda(L_x, L_y) = \frac{\left(\zeta^t_{xx}(L_x, L_y)\right)^{-1} + 2\left(\zeta^t_{yy}(L_x, L_y)\right)^{-1}}{3\left(3\pi\eta L_y\right)^{-1}} \text{ .} \tag{4.40}$$

Fig. 4.3a shows the diffusion coefficients, i.e., the inverse friction coefficients, relative to the spherical case for an ellipsoid with constant diameter $L_y = L_z$ and varying length along the symmetry axis L_x (compare the illustration of the basic geometry in Fig. 4.3b). Particularly, the plotted values are defined as:

$$\tilde{D}^i_j(L_x, L_y) = \frac{\zeta^i_j(L_y, L_y)}{\zeta^i_j(L_x, L_y)} \text{ .} \tag{4.41}$$

With this definition the relative mobility measure can be written as $\lambda = (\tilde{D}^t_{xx} + 2\tilde{D}^t_{yy})/3$. All diffusion coefficients except $\tilde{D}^r_{yy}$ do only sublinearly depend on the aspect ratio $\xi = L_x/L_y$.

In combination with the findings from the last section we are now able to implement the numerical solution of a Langevin equation as introduced in chapter 2 especially for ellipsoids *including* the crossover from anisotropic to isotropic diffusion by using $\mathbb{D}(\Delta t)$ instead of a constant $\mathbb{D}$ for all time steps Δt in the adaptive time step approach.

Collision detection

In our simulation we assume the particles to be rigid and thus use a hard-sphere-type potential, which is infinite for an overlap of two ellipsoid particles, and zero otherwise. While the detection of collisions between spheres is trivial, it is a hard problem to find an analytic overlap criterion for a pair of ellipsoids. However, suitable algorithms for this purpose have been developed in the field of hard body fluids [8, 9]. The principal approach for checking ellipsoid collisions is, e.g., well described in [7]. Typically, the starting point is the characteristic equation of an ellipsoid located at the origin and with the semi axis a, b and c oriented in parallel to the coordinate axis:

$$0 = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = \mathbf{X}\bar{\mathbb{A}}\mathbf{X}^{\mathsf{T}} \,, \tag{4.42}$$

$$\bar{\mathbb{A}} = \begin{pmatrix} 1/a^2 & 0 & 0 & 0 \\ 0 & 1/b^2 & 0 & 0 \\ 0 & 0 & 1/c^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \,, \qquad \mathbf{X} = (x,\, y,\, z,\, 1) \,. \tag{4.43}$$

This can be generalized by additionally considering a transformation $\mathbb{T}$ about an arbitrary vector (x_0, y_0, z_0) and a general rotation $\mathbb{R}$ according to:

$$0 = \mathbf{X}\mathbb{T}\mathbb{R}\bar{\mathbb{A}}\mathbb{R}^{\mathsf{T}}\mathbb{T}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}} \,, \tag{4.44}$$

$$\mathbb{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -x_0 & -y_0 & -z_0 & 1 \end{pmatrix} \,, \qquad \mathbb{R} = \begin{pmatrix} r_{xx} & r_{xy} & r_{xz} & 0 \\ r_{yx} & r_{yy} & r_{yz} & 0 \\ r_{zx} & r_{zy} & r_{zz} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \,. \tag{4.45}$$

In the following, we assume that an ellipsoid $\mathcal{A}$ is described by a generalized matrix $\mathbb{A} = \mathbb{T}\mathbb{R}\bar{\mathbb{A}}\mathbb{R}^{\mathsf{T}}\mathbb{T}^{\mathsf{T}}$. According to Ref. [158] the characteristic equation of two ellipsoids $\mathcal{A}$ and $\mathcal{B}$ is:

$$f(\lambda) = \mathsf{Det}\left[\lambda\mathbb{A} + \mathbb{B}\right] = 0 \,. \tag{4.46}$$

Because the determinant is a multiplicative map and $\mathsf{Det}[\mathbb{A}] \neq 0$ in general, the roots of $f(\lambda)$ are equivalently found as eigenvalues of the matrix $(-\mathbb{A}^{-1}\mathbb{B})$:

$$\begin{aligned} & 0 = \mathsf{Det}\left[\lambda\mathbb{A} + \mathbb{A}\mathbb{A}^{-1}\mathbb{B}\right] = \mathsf{Det}\left[\mathbb{A}\left(\lambda\mathbb{I} + \mathbb{A}^{-1}\mathbb{B}\right)\right] = \mathsf{Det}\left[\mathbb{A}\right] \cdot \mathsf{Det}\left[\lambda\mathbb{I} + \mathbb{A}^{-1}\mathbb{B}\right] \,, \\ \Rightarrow \quad & 0 = \mathsf{Det}\left[\left(-\mathbb{A}^{-1}\mathbb{B}\right) - \lambda\mathbb{I}\right] \,. \end{aligned} \tag{4.47}$$

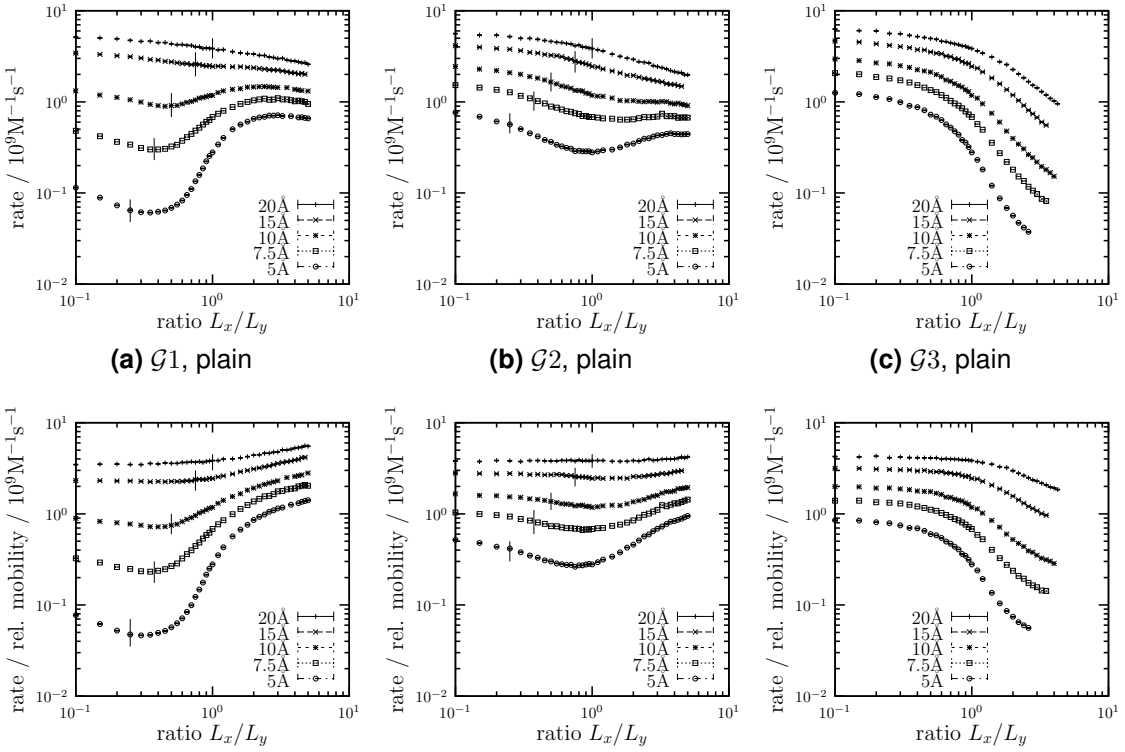
It can be shown that $\mathcal{A}$ and $\mathcal{B}$ are separated, if $(-\mathbb{A}^{-1}\mathbb{B})$ has two distinct real and positive eigenvalues. They are touching, if it has one real and positive eigenvalue of multiplicity two. As soon as there is an overlap of the two ellipsoids, the two former positive eigenvalues gain a finite imaginary component. In Ref. [158], a proof for this is given, assuming that the two ellipsoids have the same orientation (i.e. $\mathbb{R} = \mathbb{I}$). However, the conditions stated above also hold, if this constraint is not fulfilled.

The eigenvalue problem of the 4×4 matrix $(-\mathbb{A}^{-1}\mathbb{B})$ leads to a quartic equation, which can be analytically solved with, e.g., Ferrari's method. Therefore, it is in principle possible to derive a more direct criterion for the collision detection, which does not require the calculation of the eigenvalues. Such a criterion has been derived by Vieillard-Baron [152, 153]. The necessary calculations are still rather costly. Therefore, Allen et al. [8] suggested to also take into account the less costly and necessary but not sufficient condition derived by Perram et al. [108, 109]. In Ref. [10], Fortran pseudo code is given for both of these criteria. Here, we have implemented the general procedure described above in C using the CLAPACK package [13] for the determination of the eigenvalues.

4.3 Influence of anisotropic shape on encounter rate

The ellipsoid is a geometric body with a well defined anisotropy and provides both an efficient methodology for collision detection and analytic expressions for the diffusion matrix. To observe effects of shape anisotropy in analogy to the studies in chapter 3 we perform similar simulations with a pair of molecular sized ellipsoids in a periodic boundary box. The diameter in y and z direction is fixed to $L_y = L_z = 40\text{Å}$. The diameter along the x axis L_x is varied and defined by the aspect ratio $\xi = L_x/L_y$. The size of the periodic boundary box models a concentration of $c = 200\text{nM}$. Each of the model particles has a spherical reactive patch on its surface. An overlap of these reactive patches is considered as an encounter in our model. We start simulations at random initial positions and orientations and measure the time until the two model particles reach an encounter state for the first time, i.e., until the first overlap of their reactive patches occurs. This time is the first passage time (FPT) to encounter.

Besides ξ the location of the patches relative to the model particles is a possible source for varying encounter behavior. The latter is described, e.g., by altitude ψ_z and azimuth ψ_x with respect to the symmetry axis of the ellipsoid $\mathbf{e}_x$, assuming that the center of the patch is always at the surface. As in chapter 3 we average over all initial conditions. Therefore, different ψ_x are indistinguishable due to the rotational symmetry. In this case the only important parameter to define the patch position is ψ_z (compare Fig. 4.3b). A series of simulations with varying aspect ratios ξ was carried out. First, only two distinct choices for the altitude were considered, particularly $\psi_z \in \{0, \pi/2\}$. Therefore, three pairs of patch geometries are possible: The patches on both particles lie on the respective symmetry axis $\psi_z^1 = \psi_z^2 = 0$ ($\mathcal{G}1$) and one or both are aligned perpendicular to the symmetry axis, $\psi_z^1 = 0$, $\psi_z^2 = \pi/2$ ($\mathcal{G}2$) or $\psi_z^1 = \psi_z^2 = \pi/2$ ($\mathcal{G}3$) respectively.



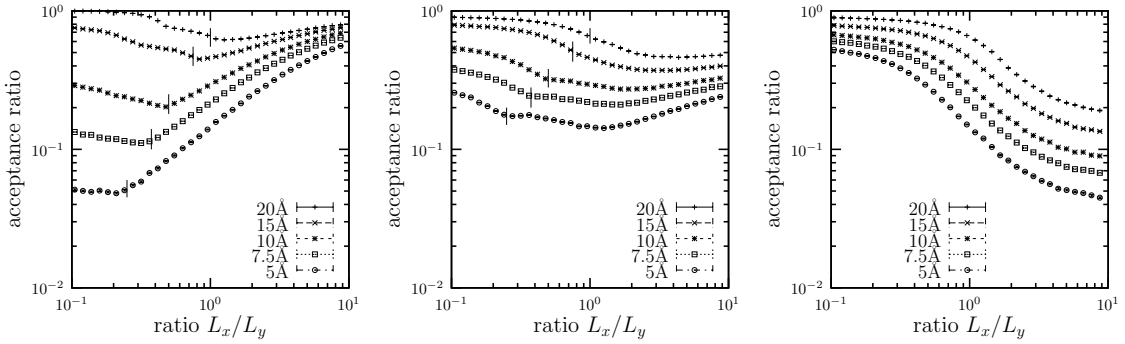
(a) $\mathcal{G}1$, plain **(b)** $\mathcal{G}2$, plain **(c)** $\mathcal{G}3$, plain

(d) $\mathcal{G}1$, mobility correction **(e)** $\mathcal{G}2$, mobility correction **(f)** $\mathcal{G}3$, mobility correction

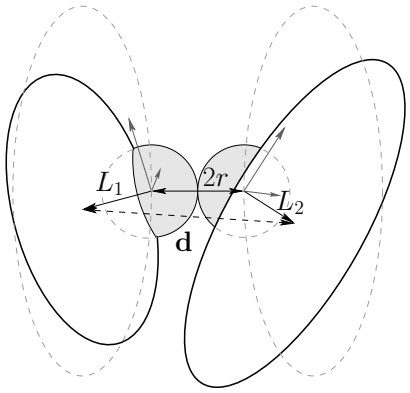
Figure 4.4: (a), (b), (c) Pure encounter rates as obtained by simulations for different patch sizes (see key in the plots). (d), (e), (f) shows the same data with a mobility correction relative to the spherical case with the aspect ratio $\xi = 1$. The patches are located according to $\mathcal{G}1$ in (a), (d), $\mathcal{G}2$ in (b), (e) and $\mathcal{G}3$ in (c), (f). The vertical lines denote the value of ξ below which erroneous encounters are counted for $\mathcal{G}1$ and $\mathcal{G}2$.

Figs. 4.4a–4.4c show the encounter rates as obtained by the simulations. An important remark has to be made regarding $\mathcal{G}1$ and $\mathcal{G}2$. Let r denote the radius of the reactive patch. As the patches are spherical and the overlap is checked by comparing their center-to-center distance with the sum of the patch radii $2r$, erroneous "back side" bindings are possible for small aspect ratios $\xi < 2r/L_y$ if at least one patch is located on the symmetry axis ($\psi_z = 0$). Therefore, there is a erroneous increase in the accessible configuration space for $\xi < 2r/L_y$ and thus the measurements are not reasonable anymore. Detecting these erroneous encounters in the simulation is hard. It requires to check whether the overlapping subvolume of the reactive patches intersect with the subvolume in front of the ellipsoid confined by the ellipsoidal surface. Therefore, we reject the complicated classification of correct and erroneous encounters. We rather denote the respective threshold by a vertical line in Figs. 4.4a and 4.4b. Only the part of the plots to the right of these lines is physically reasonable. As the L_y is constant and we do only consider patch radii $r < L_y$ no erroneous encounters are possible in case $\mathcal{G}3$.

Furthermore, by varying L_x and leaving $L_y = L_z$ constant, the effective size of the particle is altered. This will affect the overall mobility. Particularly, for large ξ the particle is effectively larger and thus less mobile than for small ξ. This will lower the encounter rate. Indeed, all plots show a systematic decrease of the



(a) $\mathcal{G}1$, steric acceptance **(b)** $\mathcal{G}2$, steric acceptance **(c)** $\mathcal{G}3$, steric acceptance



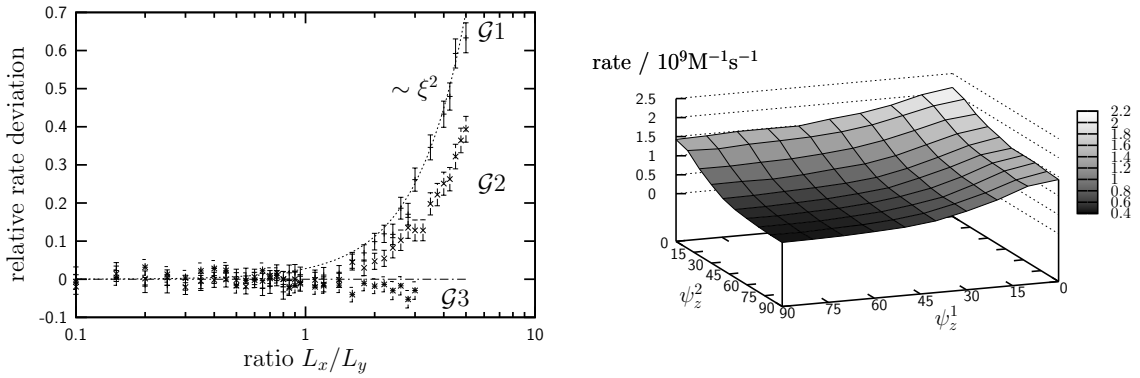
(d)

Figure 4.5: (a)–(c) Statistical estimate of the fraction of configuration space available for encounter for the patch positions and sizes as considered in Fig. 4.4. Again, the vertical lines denote the threshold below which erroneous encounters occur (see explanation in the text). (d) Illustration of the setup used to obtain the data in (a)–(c), here shown exemplary for $\mathcal{G}1$.

encounter rate with increasing ξ besides their specific behavior. Recalling the Smoluchowski equation Eq. (1.1) the encounter rate is expected to be linearly dependent on the mobility in the respective simple framework. Therefore dividing the results shown in Figs. 4.4a–4.4c by $\lambda(\xi)$ as defined in Eq. (4.40) approximately corrects the error due to the varying size of the model particles. The corrected data is shown in Figs. 4.4d–4.4f. From the plots it is apparent that the encounter rate is enhanced for large ξ in $\mathcal{G}1$ and for small ξ in $\mathcal{G}3$. As the two cases are qualitatively different the behavior is not symmetric in a strict sense. However, as $\mathcal{G}2$ can be considered as a mixture between $\mathcal{G}1$ and $\mathcal{G}3$ it is reasonable that it shows a quasi-symmetric behavior around the intermediate regime of $\xi \approx 1$.

Contribution of steric effects

Before investigating the reasons for the observed effects another important systematic difference between ellipsoids with different ratios ξ has to be noted: As the geometry of the ellipsoid is changing with ξ, the exposed volume fraction f_V of the reactive patches not covered by the steric particle is changing. For example, if the patch is located on the front edge of the ellipsoid along the symmetry axis ($\mathcal{G}1$), for large ξ the fraction will approach $f_V = 1$, while for small ξ the particle has a disc like shape and thus ideally covers half of the patch $f_V = 0.5$, neglecting the erroneous encounter problem. Regarding $\mathcal{G}3$ the situation is nearly opposite. Large fractions f_V reflect a larger space of possible encounter configurations, which will have a positive influence on the encounter rate. Unfortunately, the subspace of accessible encounter configurations of two ellipsoids is a complex



(a) **(b)**

Figure 4.6: (a) Relative deviation of the encounter rates assuming isotropic motion at all times from the original data (compare Fig. 4.4). The reactive patches have a radius of $r = 10\text{Å}$. The dotted line indicates a supposed power-law. For the sake of clearness the chain dotted line hints correspondence of the data, i.e. $\Delta_{\text{iso}} = 0$. (b) Encounter rates for two ellipsoids with aspect ratio $\xi = 2.4$, encounter patch radius $r = 10\text{Å}$ and varying patch positions determined by ψ_z^1 and ψ_z^2, scaled in degrees here, as visualized in Fig. 4.3b.

six-dimensional region because of three relative degrees of freedom for translation and rotation respectively. To estimate this steric effect on the encounter rate we study the fraction of non-overlapping ellipsoid configurations with touching reactive patches. A scheme of the setup is shown in Fig. 4.5d. Particularly, the centers of the patches with radius r are placed at the distance $2r$ and both ellipsoids are then randomly rotated around the center of their respective patch. In Fig. 4.5d the resulting distance vector $\mathbf{d}$ is illustrated, which is the sum of the patch distance vector (with absolute value $2r$) and the two vectors from the respective patch centers to the rotated ellipsoid centers with length L_1 and L_2 (for $\mathcal{G}1$ as shown in Fig. 4.5d it is $L_1 = L_2 = L_x$). After obtaining a random encounter configuration we check whether the ellipsoids overlap. To gain appropriate statistics we repeat this procedure 10^5 times and calculate the fraction of non-overlapping configurations. The results for the parameters corresponding to Fig. 4.4 are shown in Figs. 4.5a–4.5c. As the method of random rotations does again allow for erroneous encounters at small $\xi < \xi_c$, the reasonable data is bounded by vertical lines in Figs. 4.5a and 4.5b. The plots show that the qualitative features of the encounter rates are well reproduced by the steric constraints. The latter correspond to the thoughts about the accessible patch area f_V. This finally leads to the conclusion that the main reason for the changes in the encounter rate in the preceeding study is not the altered hydrodynamic behavior of the ellipsoids but the steric hindering of encounters due to the changing geometry. Note that hydrodynamic two-body interactions are neglected as usual. Actually it is doubtful that continuum approaches work for the small distances considered here.

Direct comparison with isotropic diffusion

The effect of the different hydrodynamic properties for ellipsoids of varying aspect ratio ξ is screened by the dominating steric effects in our approach. Therefore, we

want to compare our simulations with explicit anisotropic diffusion to simulations, where we do *not* account for the anisotropy in the diffusion matrix. Particularly, we use an isotropic diffusion matrix with the average translational and rotational diffusion coefficients $D^t = (D^t_{xx} + D^t_{yy} + D^t_{zz})/3$, which is equal to the isotropic limit of Eq. (4.30) for large time steps $t \to \infty$, and $D^r = (D^r_{xx} + D^r_{yy} + D^r_{zz})/3$ respectively. Fig. 4.6a shows the relative deviation of the encounter rates when performing the same simulations which were used to produce the data shown in Fig. 4.4 with an isotropic diffusion matrix. Particularly, the data in the plot is:

$$\Delta_{\text{iso}} = k_{\text{isotropic}}/k_{\text{anisotropic}} - 1 \; . \tag{4.48}$$

Interestingly, if the patches are aligned along a radial axis and thus their position is independent of ξ, i.e. in case $\mathcal{G}3$, there is no significant deviation from the original results. However, considering $\mathcal{G}1$, where the patches are located at the front end of the ellipsoid along the symmetry axis, the (artificial) isotropic encounter rate is clearly higher. This effect can also be observed in the "mixed" case $\mathcal{G}2$, decreased by roughly a factor 2. This is reasonable as here only one of the two encountering ellipsoids has its patch at $\psi_z = 0$. These findings show that the anisotropic diffusion of elongated ellipsoids leads to a decrease of the encounter rates. Unfortunately, the large statistical errors do not allow for a clear statement about the correlation of Δ_{iso} and ξ, although the data fit well a $\Delta_{\text{iso}} \sim \xi^2$ power-law as indicated by the dotted line in the plot.

Finally, Fig. 4.6b shows a more detailed study regarding the possible positions of the patches, i.e., for more pairs of ψ^1_z, ψ^2_z as the three that have been considered so far. The plot shows the rates obtained by simulations with $\xi = 2.4$ and patch radius $r = 10\text{Å}$. Simulations are only performed for $\psi^1_z \geq \psi^2_z$, as due to averaging over all initial conditions the change of $\psi^1_z \leftrightarrow \psi^2_z$ does not alter the system. The data shows a flat crossover between the already considered cases $\mathcal{G}1$, $\mathcal{G}2$ and $\mathcal{G}3$. That is, more general patch alignments do not lead to any complex behavior. Here, the anisotropic diffusion matrix is taken into account. However, in analogy to Fig. 4.6a a similar analysis reveals, that the deviation of the encounter rates obtained with isotropic diffusion are largest for $\mathcal{G}1$.

4.4 Discussion

On large time scales, diffusion in an isotropic environment becomes isotropic. Diffusion of anisotropic objects is anisotropic on small time and length scales if rotational relaxation has not yet occurred. The memory of the initial orientation is important for anisotropy and decays exponentially. The analytic treatment of this problem reveals that the dominating time scale for this process is $1/(6D^r) = 1/(2c)$ (compare Eq. (4.25) and Eq. (4.31)) where D^r denotes the average rotational diffusion coefficient. The crossover from anisotropic to isotropic diffusion actually has two time scales which increasingly deviate with increasing anisotropy. In three dimensions alternating rotations around two axis are sufficient to reach any orientation. This gets apparent in the example shown in Fig.

4.2, where rotational diffusion about axis $\mathbf{e}_1$ does not occur $D_1^\theta = 0$. Therefore the two time scales for the crossover ($c \pm d$, see Eq. (4.31)) are defined by two effective rotation speeds mixing different orientation spaces, particularly the $\mathbf{e}_1$-$\mathbf{e}_2$ plane and the $\mathbf{e}_{1/2}$-$\mathbf{e}_3$ plane.

The effects of anisotropic diffusion on encounter rates is studied using the example of ellipsoids with reactive patches. The magnitude of the diameter along the symmetry axis L_x is changed by factors of $\xi = 0.1 \ldots 5$ with respect to the fixed diameter L_y. The appropriate mobility is evaluated according to analytic expressions known from the literature. By simulations we investigate pairwise encounter rates of ellipsoids with three different combinations of patch locations ($\mathcal{G}1$–$\mathcal{G}3$). If both patches are located on the symmetry axis ($\mathcal{G}1$), the obtained encounter rates are much lower for oblate-shaped ellipsoids. If both reactive patches are located perpendicular to the symmetry axis ($\mathcal{G}3$), encounter rates strongly decrease in the prolate regime. However, by varying L_x not only the degree of anisotropy of the ellipsoid with respect to diffusion changes. Firstly, the overall mobility of particles increases when ξ is decreased. We correct this effect by a relative mobility factor λ. Secondly, the shape of the ellipsoid also varies according to ξ which leads to differences in the accessible volume of the reactive patches due to excluded volume. We numerically evaluate an approximate measure for the accessibility of the patches f_V. Although the particular impact of f_V on the encounter rates is unknown, the qualitative behavior is very similar. This indicates that steric effects are the main reason for the changes in the encounter rates according to our approach. Finally, we compare our results with simulations in which we do not consider the hydrodynamic anisotropy. We find agreement irrespective of ξ if the patches are located perpendicular to the symmetry axis ($\mathcal{G}3$). Otherwise, a slow-down of encounter proportional to ξ^2 is obtained for prolates. Particularly, if both patches are at the tip of the ellipsoid ($\mathcal{G}1$) for $\xi = 5$, the relative deviation with neglecting hydrodynamic anisotropy is $\Delta_{\mathrm{iso}} \approx 0.7$.

General thermodynamic effects of suspensions of hard ellipsoids or spherocylinders at higher densities have been studied by other approaches from statistical mechanics like density functional theory. For example, Roth et al. [120] observed the depletion-induced torques and forces on spherocylinders close to a wall. Such might have an impact on the binding characteristics of, e.g., anisotropic molecules in appropriate environments. Chelakkot et al. [35] observed phase transitions of dispersions of spherocylinders with sticky ends in dependence of thermodynamic parameters. Note, however, that the focus of this chapter is on the dynamics of specific encounter events.

An interesting question that has not been investigated yet is whether the impact of hydrodynamic anisotropy regarding *re-encounter* is stronger. Because prolates have a lower overall rotational diffusion coefficient one might assume that they are much more likely to return to an encounter even after having lost contact due to diffusion. Possibly, this could accelerate bond formation in cases of protein-protein interaction where a large number of encounters is required to enable bond formation. In such situations, quick re-encounter could increase protein-protein association in analogy to the question of contact numbers in chapter 3.

Chapter 5

Dynamics of clusters

The overriding aim of this work is to use concepts from stochastic dynamics to develop an efficient yet realistic model for protein complexes which can be upscaled to large systems. We therefore do not consider atomistic details and sophisticated interaction potentials but rather focus on the concept of reactive patches and encounter complex. In this chapter we generalize our model of bimolecular encounter (see chapter 3) to address the assembly of more than two proteins. We start with three homogeneously encountering particles on a line. This problem is more complex to solve analytically than the two-particle problem treated in chapter 2. We derive and solve an exact equation for this mean first passage time problem. Next we consider complex formation in three dimensions. Particularly, we choose a model system of particles forming equilateral three-particle clusters. We take special care to derive appropriate rules for the alignment after bond formation. The extraction of rate data for a stochastic network view on the complex dynamics from simulations is explained by assuming that all processes in the network are of Poisson type. However, transition time histograms of our model system show that this assumption is not valid. This means that the assembly network can not be evaluated by a Master equation approach in general. Average populations and fraction time series show that transport steps have a strong impact on the system which qualitatively differs from expectations on the basis of the reaction steps only. Finally, the self-assembly dynamics of a T1 virus capsid structure is considered as a first application to larger systems. The results for varying unbinding rates agree well with recent findings by similar model approaches reported in the literature.

5.1 Cluster of three homogeneously binding particles

We begin the considerations regarding the dynamics of multi-particle complexes by investigating the association dynamics of three homogeneously binding particles in analogy to Sect. 2.2.2. That is, we disregard the concept of having reactive patches at the surface of the particles and assume that they are fully covered with ligands and receptors, respectively. An interesting but non-trivial question is how

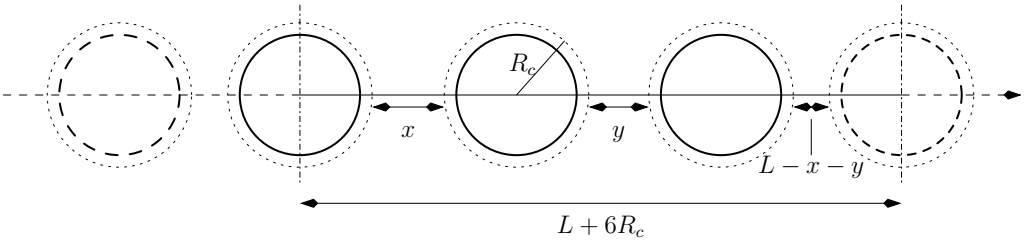


Figure 5.1: Illustration of the problem of three particles in a box with periodic boundary conditions in one dimension. Assuming the boundary to be at the center of one of the spheres, any state of the system can be fully described by the box length and two independent variables x and y. In this picture, L is the "free" length of the box, i.e. after subtracting the finite sizes of the particles.

long three particles need to successively bind to form one cluster. As usual we assume periodic boundary conditions. This problem can be best addressed in one dimension. Fig. 5.1 shows a schematic illustration of the setup. The boundary is placed at the center of one of the spheres. L is the sum of the distances between the absorbing shells, so that $L + 6R_c$ is the total size of the box. Therefore, it is possible to describe a certain state of the system with the two distances x and y. The third is then determined by the boundary conditions $L - x - y$. In a system with instantaneous binding after encounter and without dissociation, the process of cluster formation consists of two parts. First, two of the three freely diffusing particles have to bind to each other. Afterwards, a process as treated in Sect. 2.2.2 (but with different D_t for the two particles) then leads to the final clustering.

The first process has to be discussed in more detail. The corresponding mean first passage time will be denoted as $T^{(1)}$. The problem of a particle binding to either its left or its right neighbor in one dimension is similar to the problem discussed in Sect. 2.2.2. However, the difference is that in Sect. 2.2.2 the distance between absorbing boundaries is constant. In the present case, the boundaries can fluctuate. In order to derive a differential equation for the calculation of $T^{(1)}$ we consider particle movement on a lattice with step size δ and stepping time τ and obtain a continuous description in the limit of infinitesimally small steps. In each time step each of the three particles either moves to the left (-) or the right (+). This gives 8 different possibilities of movement in the whole system with equal probability for each step and respective changes of x and y: +++ ($x \to x$, $y \to y$), ++- ($x \to x$, $y \to y - 2\delta$), ... Thus, the MFPT can be expressed as:

$$T^{(1)}(x,y) = \tau + \frac{1}{8}\Big(T^{(1)}(x+2\delta,y) + T^{(1)}(x-2\delta,y) + T^{(1)}(x,y+2\delta) + T^{(1)}(x,y-2\delta) + T^{(1)}(x+2\delta,y-2\delta) + T^{(1)}(x-2\delta,y+2\delta) + 2T^{(1)}(x,y)\Big) \,. \quad (5.1)$$

By complementing the sum with pairwise compensating terms and after regroup-

ing, one ends up with:

$$\begin{aligned}-8\tau = &2\left(T^{(1)}(x+2\delta,y)+T^{(1)}(x-2\delta,y)-2T^{(1)}(x,y)\right)\\ &+2\left(T^{(1)}(x,y+2\delta)+T^{(1)}(x,y-2\delta)-2T^{(1)}(x,y)\right)\\ &-\left(T^{(1)}(x+2\delta,y)-T^{(1)}(x,y)-T^{(1)}(x+2\delta,y-2\delta)+T^{(1)}(x,y-2\delta)\right)\\ &-\left(T^{(1)}(x,y+2\delta)-T^{(1)}(x-2\delta,y+2\delta)-T^{(1)}(x,y)+T^{(1)}(x-2\delta,y)\right)\ .\end{aligned} \tag{5.2}$$

Dividing both sides of the equation by $(2\delta)^2$ and in the limit of vanishing δ and τ with $D=\delta^2/\tau$, the final differential equation to describe $T^{(1)}$ is:

$$-\frac{2}{D}=2\left[\frac{\partial^2}{\partial x^2}T^{(1)}(x,y)+\frac{\partial^2}{\partial y^2}T^{(1)}(x,y)-\frac{\partial^2}{\partial x\partial y}T^{(1)}(x,y)\right]\ . \tag{5.3}$$

We now derive the solution to Eq. (5.3) in a systematic manner. A specific solution to the inhomogeneity of Eq. (5.3) is:

$$T^{(1)}(x,y)=-\frac{1}{2D}x^2\ . \tag{5.4}$$

The homogeneous part of Eq. (5.3) is solved by any linear combination of terms, which are either the imaginary or the real part of:

$$f_n(x,y)=\left(\frac{1}{2}\left(1+\imath\sqrt{3}\right)x+y\right)^n\ . \tag{5.5}$$

The boundary conditions for the considered case are:

$$T^{(1)}(x,y)\big|_{x=0}=T^{(1)}(x,y)\big|_{y=0}=T^{(1)}(x,y)\big|_{x+y=L}=0\ . \tag{5.6}$$

The first two conditions $T^{(1)}(0,y)=T^{(1)}(x,0)=0$ imply that there can be no terms in $T^{(1)}$ which depend exclusively on x or on y. Since the real part of f_n always contains at least an y^n summand, $T^{(1)}$ can only consist of the respective imaginary parts, which are for $n=\{1,2,3\}$:

$$\mathrm{Im}\left\{f_1(x,y)\right\}=\frac{\sqrt{3}}{2}x\ , \tag{5.7}$$

$$\mathrm{Im}\left\{f_2(x,y)\right\}=\frac{\sqrt{3}}{2}x^2+\sqrt{3}xy\ , \tag{5.8}$$

$$\mathrm{Im}\left\{f_3(x,y)\right\}=\frac{3}{2}\sqrt{3}(x^2y+xy^2)\ . \tag{5.9}$$

$\mathrm{Im}\{f_1\}$ does not fulfill the boundary condition $T^{(1)}(x,0)=0$. Also the specific inhomogeneous solution Eq. (5.4) violates this condition. However, this can be corrected with the respective counterpart from $\mathrm{Im}\{f_2\}$. The result is:

$$T'(x,y)=-\frac{1}{2D}x^2+\frac{1}{\sqrt{3}D}\left(\frac{\sqrt{3}}{2}x^2+\sqrt{3}xy\right)=Dxy\ . \tag{5.10}$$

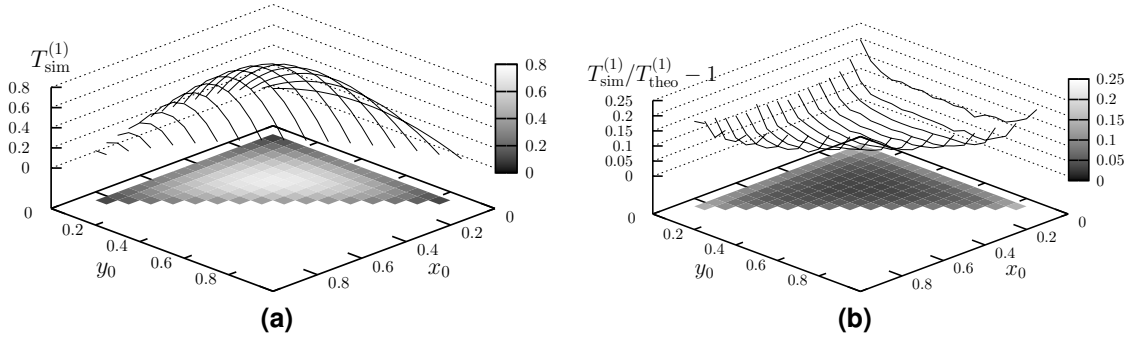


Figure 5.2: Mean time to first encounter of any two particles of the system described in Fig. 5.1 in dependency of the initial state (x_0, y_0). We use the parameter set $\eta_0 = R = 1$, $R_c = 1.1$ and time step $\tau = 0.001$. The box length is chosen such that $L = 1$. The simulation data are shown in (a) and the relative deviation from the analytic calculation according to Eq. (5.11) is plotted in (b).

However, T' still does not satisfy the third boundary condition $T^{(1)}(x, L-x) = 0$. Therefore, a term of form $\mathrm{Im}\{f_3\}$ is required to obtain the final result:

$$T^{(1)}(x,y) = \frac{1}{D}xy - \frac{1}{DL}(x^2y + xy^2) = \frac{1}{DL}xy\,(L - (x+y)) \ . \tag{5.11}$$

It is easy to check that this expression is indeed a solution of Eq. (5.3). The final expression is very similar to the one obtained for the case of two particles in Eq. (2.32). $T^{(1)}$ is linear in each of the possible initial distances and the correct dimension is preserved by the factor $1/L$, as L is the only length scale in the system. We perform corresponding simulations with three particles on a line. The simulation data is shown in Fig. 5.2a. The analytic solution looks very similar and the relative deviation is shown in Fig. 5.2b. Obviously the deviation is largest at the boundaries of the initial configuration space (up to $\sim 20\%$), while in the intermediate region it is much smaller ($\sim 1\%$). This is similar to the finding in Fig. 2.7a, which can be explained by the systematic error due to the finite time steps as explained in Sect. 2.2. It is reasonable that the relative error $\mathcal{O}(\sqrt{\Delta t}/T)$ is largest for small MFPTs close to the boundaries because the absolute systematic error does not depend on the initial configuration.

Coming back to the complete clustering process, one might expect that during the first step, which is the diffusion until two particles encounter and immediately bind, the memory of the system with respect to the initial configuration should get lost. Particularly, this would mean that all initial distances between the two-particle cluster and the still free particle, would have an equal probability. In this case, one could average over all possible distances and would then get an average first passage time for the total binding, which could be added to the first part. However, Fig. 5.3a shows that also the second part of the clustering process depends on the initial configuration. The graphs for $T(x_0, y_0)$ have the parabolic form which also was observed for $T^{(1)}$, but the quantitative behavior is changed. While e.g. for small x_0 the first encounter is always reached quickly, the second one takes much longer, especially for intermediate values of y_0. On the other

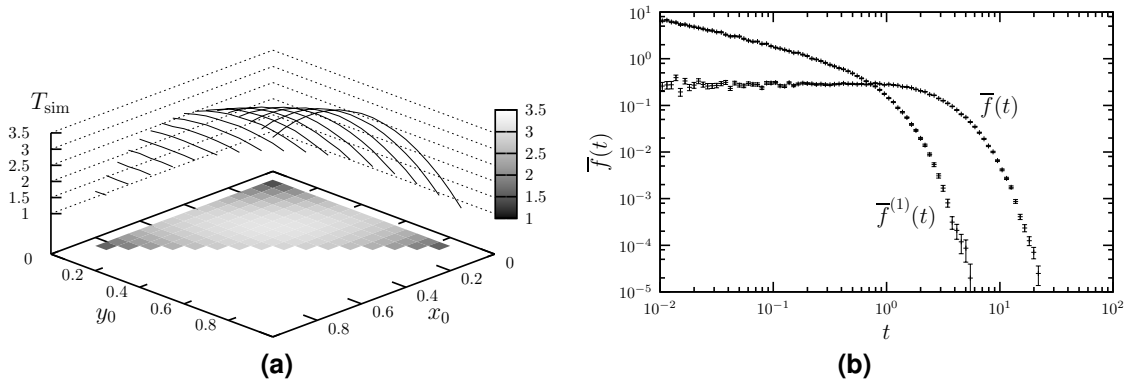


Figure 5.3: (a) Mean first time to a bound three-particle cluster in the system described as in Fig. 5.2a for the same set of initial conditions. (b) Double-logarithmic plot of first passage time distributions of the first ($\overline{f}^{(1)}$) and complete binding ($\overline{f}$) in the considered system (see caption of Fig. 5.2a) averaged over possible initial contributions.

hand, for intermediate values of x_0, the differences in $T^{(1)}$ for certain y_0 are less distinct for the over-all T. The whole problem is again symmetric in x_0 and y_0.

In Fig. 5.3b, the FPT distributions for the first binding and the complete cluster formation averaged over the initial conditions are shown. Astonishingly, $\overline{f}^{(1)}$ is also $\sim 1/\sqrt{t}$ at small t and exponentially decays for high t. This is qualitatively equal to the case discussed in Sect. 2.2.2. Furthermore, $\overline{f}$ looks very similar to a Poisson distribution, as it shows the characteristic plateau for small t. The second encounter is in principle similar to the encounter of two particles discussed in chapter 2. If the initial configurations of this process after the first encounter would be equally distributed one would expect $\overline{f}^{(2)}$ to be qualitatively equal to $\overline{f}^{(1)}$ and $\overline{f}$ could be generated by a convolution. However, the convolution of two processes with distributions like $\overline{f}^{(1)}$ cannot produce a Poisson distribution. Therefore, the form of $\overline{f}$ is another hint that the second encounter process is biased by the particular realization of the first one.

5.2 Three-particle cluster with distinct geometry

The preceding section dealt with homogeneously binding particles. In line with our investigations of bimolecular binding we now extend our cluster model by the concept of finite-sized reactive patches. In principle, we follow the spirit of the model presented in chapter 3 to study clusters of more than two particles. However, we now have to proceed beyond the first encounter and must specify rules for the cluster structure.

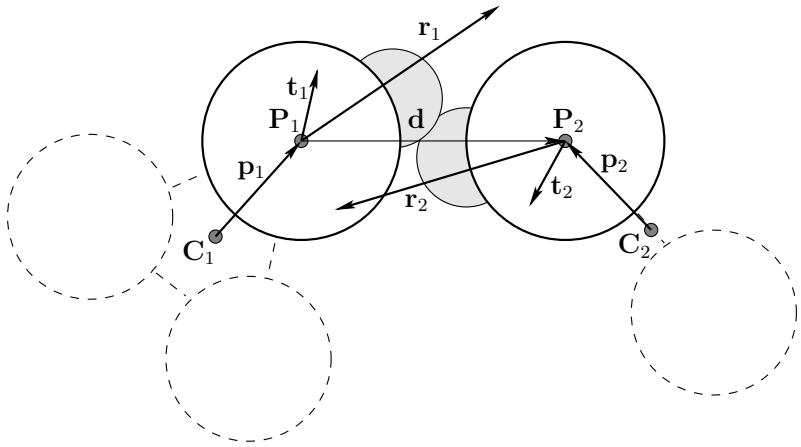


Figure 5.4: Generic illustration of two clusters forming an encounter. The vectors $\mathbf{r}_{1/2}$ determine the desired final relative positions of the centers of the binding proteins $\mathbf{P}_{1/2}$. The two vectors $\mathbf{t}_{1/2}$ define the orientation with respect to the final linking vector $\mathbf{d}''$. The necessary transformations will also affect the center positions of the two clusters $\mathbf{C}_{1/2}$.

5.2.1 Aligning bound model proteins in three dimensions

To build up larger clusters with a distinct geometry as given, e.g., by a specific structure contained in the protein data base (PDB), it is necessary that two binding proteins end up in a particularly specified relative position and orientation. In the real biologic situation this is achieved by short-ranged interactions between the binding interfaces, which clearly define a small window in configuration space as energetically favorable in analogy to the key-lock binding hypothesis. However, this conformation is still subject to thermal fluctuations. It is even possible that two bound proteins alternate between several conformations. As we do not want to consider submolecular fluctuations in our model and our particles are meant as an averaged representation of the biological proteins, we neglect such conformational fluctuations between bound proteins.

If two proteins contained in different clusters bind in the model, the respective clusters are combined. As each protein bond implies a well-defined relative alignment of the two partners and thus the structure of a cluster is fixed, the alignment requirements for the newly evolving protein bond can only be fulfilled by changing the position and orientation of the whole clusters containing the binding partners. In our model, this reorientation and translation happens instantaneously. A more accurate model for the dynamics leading to the final conformation is not required for our purpose. Often, artificial potentials are used to favor a certain alignment state. This approach is not necessarily more physical and would lead to undesirable high frequency oscillations in the numerical solution due to the finite time step. To prevent errors from these oscillations a small time step has to be chosen which strongly slows down the simulation speed.

Fig. 5.4 should help to illustrate the applied procedure to realign the clusters as required to proceed with the next step after binding. The figure schematically

shows two proteins with center position $\mathbf{P}_{1/2}$ and overlapping reactive patches in light gray. Each protein is part of a small cluster, and the already bound proteins are indicated with dashed lines. The hydrodynamic center of the respective clusters, which is the point of reference for the cluster translation and rotation, is denoted with $\mathbf{C}_{1/2}$. The position of a protein relative to the center of the cluster, in which it is contained, is given by $\mathbf{p}_{1/2}$. The structure of a bond is uniquely defined by two vectors for each of the two binding proteins, both given relative to their center. First, $\mathbf{r}_{1/2}$ determines the position of the opposing protein in the bound state. However, even if the clusters would be translated and rotated such that $\mathbf{r}'_1$ points to $\mathbf{P}'_2$ and vice versa (with primes indicating *transformed* vectors), there would still be one degree of freedom left. Particularly, the two clusters could be rotated around the linking vector $\mathbf{r}'_1 = -\mathbf{r}'_2$. Therefore, a further vector $\mathbf{t}_{1/2}$ is defined for each protein, which must not be parallel or antiparallel to $\mathbf{r}_{1/2}$, respectively. That is, the projection of $\mathbf{t}'_{1/2}$ on the plane perpendicular to $\mathbf{r}'_{1/2}$ does not vanish. The equality of the two projected vectors then gives the missing condition for full determination of the bond structure.

In the simulation, the transformations are carried out in three steps. First, the clusters are rotated such that $\mathbf{r}_1$ becomes antiparallel with $\mathbf{r}_2$. A second rotation then brings the $\mathbf{t}$ vectors into correspondence. Finally, the clusters are translated such that $\mathbf{P}'_1 + \mathbf{r}'_1 = \mathbf{P}'_2$ and vice versa. The moves required by the bond formation should be kept as small as possible. Furthermore, as the combining clusters can have very different sizes, there should be some balancing of the motion. If, e.g., a small cluster is bound to a much larger one, the largest part of the required transformations should act on the smaller one. The chosen balancing is based on the diffusion properties. In particular, overall translational and a rotational diffusion coefficients are calculated for each cluster by averaging the trace of the respective parts of the diffusion matrix:

$$D_t = (\mathcal{D}^t_{xx} + \mathcal{D}^t_{yy} + \mathcal{D}^t_{zz})/3 \,, \qquad D_r = (\mathcal{D}^r_{xx} + \mathcal{D}^r_{yy} + \mathcal{D}^r_{zz})/3 \,. \tag{5.12}$$

The necessary translation distances and rotation angles are then assigned to both clusters weighted according to:

$$B^i_t = D^i_t/(D^1_t + D^2_t) \,, \qquad B^i_r = D^i_r/(D^1_r + D^2_r) \,, \tag{5.13}$$

respectively. In the following the applied transformations will be explained in more detail.

Fig. 5.5 illustrates the rotation to align $\mathbf{r}_1$ and $\mathbf{r}_2$ antiparallelly. The enclosed angle is:

$$\theta_r = \arccos\left(\frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{|\mathbf{r}_1||\mathbf{r}_2|}\right) \,, \tag{5.14}$$

and must be increased to $\theta'_r = \pi$. Thus, the rotation acts in the plane spanned by the two vectors $\mathbf{r}_{1/2}$ and the rotation axis is therefore given by:

$$\mathbf{\Omega}_r = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{|\mathbf{r}_1||\mathbf{r}_2|} \,. \tag{5.15}$$

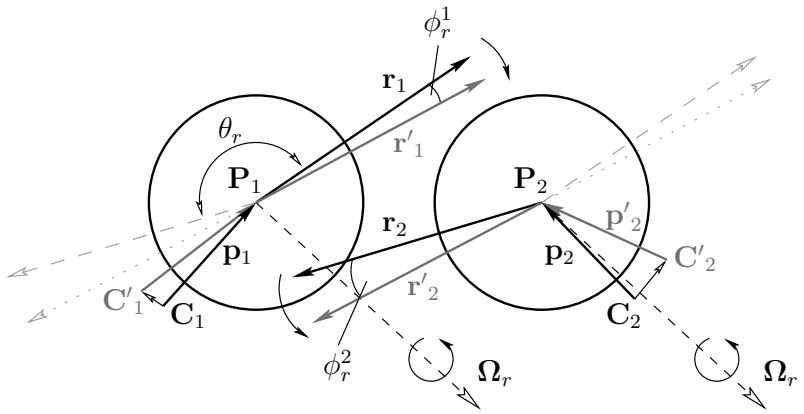


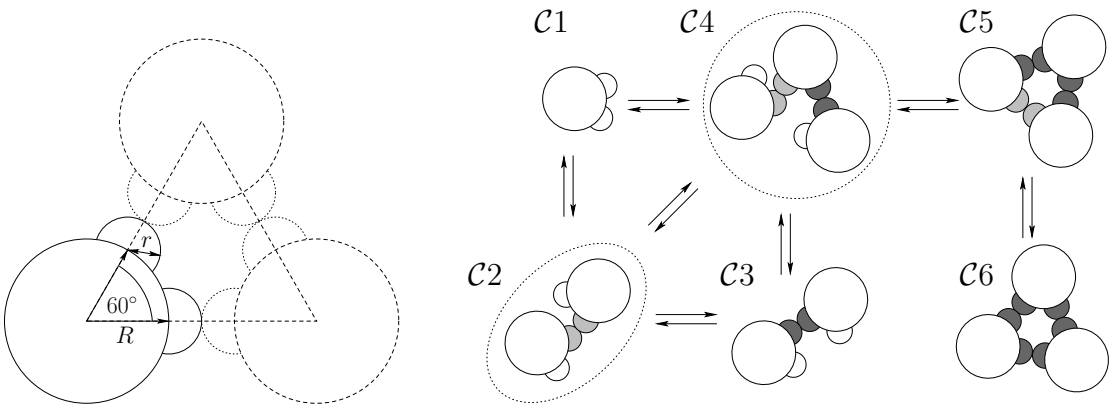
Figure 5.5: Illustration of the rotation to align the center-to-center vectors $\mathbf{r}_{1/2}$. The affected vectors are shown in black before and dark gray after the rotation. For the sake of clarity, copies of $\mathbf{r}$ from the opposing protein are hinted in light gray. The cross product of these vectors gives the rotation axis $\mathbf{\Omega}_r$, depicted as a dashed arrow. After the rotation $\mathbf{r'}_1$ and $\mathbf{r'}_2$ are antiparallel and form a straight line in the figure. $\mathbf{t}_1$ and $\mathbf{t}_2$ (compare Fig. 5.4) are not shown, although they are also subject to the rotation.

The angles of rotation according to the previously explained balancing rule are $\phi_r^1 = -(\pi - \theta_r)B_r^1$ and $\phi_r^2 = (\pi - \theta_r)B_r^2$ respectively. Note that the minus sign in the definition of ϕ_r^1 already indicates that the first cluster has to be rotated clockwise around $\mathbf{\Omega}_r$. As the cluster rotation acts with respect to the hydrodynamic centers $\mathbf{C}_{1/2}$, this could actually cause a serious translation of the proteins forming the bond, especially if both clusters are large. Therefore, in the model we try to let the bond develop close to the position, where the patch overlap occurred. Thus, for the cluster rotations the centers of the binding proteins are chosen as the point of reference. This means that $\mathbf{C}_i$ have to be translated by $\mathbf{p}_i - \mathbf{p'}_i$, where $\mathbf{p'}_i$ is the vector $\mathbf{p}_i$ rotated around $\mathbf{\Omega}_r$ by ϕ_r^i.

The axis for the second rotation must be parallel to $\mathbf{r'}_1$, as this will be the linking vector between the bound proteins (after a final translation). Therefore, the rotation vector is now $\mathbf{\Omega}_t = \mathbf{r'}_1/|\mathbf{r'}_1|$. The *projections* of $\mathbf{t'}_1$ and $\mathbf{t'}_2$ on the plane perpendicular to $\mathbf{r'}_{1/2}$ are required to be parallel. As only the directions of $\mathbf{t}_{1/2}$ are important, the projection can be circumvented by considering the cross products $\boldsymbol{\chi}_{1/2} = \mathbf{t'}_{1/2} \times \mathbf{r'}_1$, respectively. While computationally less costly, this is equivalent to the projection approach, as $\boldsymbol{\chi}_{1/2}$ will be automatically perpendicular to $\mathbf{r'}_1$ and basically tilted against the projections of $\mathbf{t'}_{1/2}$ by $\pi/2$. As the calculation of angles via $\arccos$ as already applied above only gives the enclosed angle between vectors, the direction of rotation has to be determined. This can be done by comparing the cross product $\boldsymbol{\chi}_1 \times \boldsymbol{\chi}_2$ with $\mathbf{r'}_1$. If both point into the same direction the rotation angle for the first cluster is positive and vice versa. So the overall rotation angle is

$$\theta_t = \arccos\left(\frac{\boldsymbol{\chi}_1 \cdot \boldsymbol{\chi}_2}{|\boldsymbol{\chi}_1||\boldsymbol{\chi}_2|}\right) \cdot \mathrm{sgn}((\boldsymbol{\chi}_1 \times \boldsymbol{\chi}_2) \cdot \mathbf{r'}_1) \;, \tag{5.16}$$

where sgn is the sign function. Consequently the portions for each cluster are



(a) Geometry **(b)** Substate network

Figure 5.6: (a) The geometry of the reactive patches for a simple cluster of three equal particles is straightforward. The two patches should be located with an enclosed angle of $60°$. In the bound state, the patch vectors will have to be antiparallel, i.e., the linking vector between the center of two of these particles will be parallel to its respective patch vector. (b) Any substate in the system described in (a) can be characterized as one of four complexes, including the single particle ($\mathcal{C}1$), and two encounter complexes. Here, a generic network including all possible transitions between the different substates is shown. The encounter complexes $\mathcal{C}2$ and $\mathcal{C}4$ are surrounded by a dashed line. Furthermore, the reactive patches which are in an encounter state are colored in light gray. In all complexes with more than one particle, the formed bonds are denoted by patches colored in dark gray. Note that although $\mathcal{C}5$ is already a complex of three particles, it is kinetically different from $\mathcal{C}6$, where each particle is bound to both neighbors. The substates are numbered in the order they will occur on the way from the single particle $\mathcal{C}1$ to the fully bound complex $\mathcal{C}6$.

$\phi_t^1 = \theta_t B_r^1$ and $\phi_t^2 = -\theta_t B_r^2$, respectively. As for the $\mathbf{\Omega}_r$ rotation, the position of the binding proteins should be kept fix and thus the clusters have to be translated about $\mathbf{p'}_i - \mathbf{p''}_i$, where two primes indicate the state after this *second* rotation.

While the orientations of the clusters (and especially the proteins) are now as desired, the position is still not necessarily matching the requirement $\mathbf{d''} = \mathbf{d} = \mathbf{r}_1$. $\mathbf{d}$ denotes the distance of the centers of the two binding proteins in the encounter state, i.e., before the binding transformation. As the rotations leave the protein centers $\mathbf{P}_{1/2}$ conserved it holds $\mathbf{d} = \mathbf{d''}$. Therefore, the final transformations are $(\mathbf{d} - \mathbf{r''}_1)B_t^1$ and $-(\mathbf{d} - \mathbf{r''}_1)B_t^2$ for clusters 1 and 2, respectively.

5.2.2 Network view of assembly – substates and transition dynamics

In Fig. 1.3 from the introduction we illustrate all possible substates during the formation of a three particle cluster for the case that any pair of particles has to bind at one specific pair of interaction sites. We want to study a somewhat simpler case here, where each of the model particles has two binding interfaces which can form a bond with any of the other interfaces, i.e., all particles are of the same

type (see Fig. 5.6a). As shown in Fig. 5.6b this scheme gives rise to six different substates, counting all possible partial and encounter complexes up to the fully bound complex. In principle, also an encounter complex of two two-particle clusters $\mathcal{C}3$ is possible. However, due to the excluded volume properties of the particle the bond cannot be formed and thus we do not consider it in our network. We basically consider *local* states, i.e., according to its particular configuration each particle in the simulation can be assigned to one of the six substates shown in Fig. 5.6b. A different approach would be to define a network of global *micro states* of the whole system. That is, each state reflects one particular configuration of the entity of all particles (e.g.: "all particles are separate"). However, the latter view leads to a rapid increase of the network complexity with the number of particles.

In our binary view of protein interaction, the transitions between these states can be due to either a transport or a reaction process. The reaction processes are Poisson processes with predefined rates by the design of our model. These rates have to be reproduced by the simulation. The transport is the more complicated encounter process as discussed in detail in chapter 3. We basically generalize typical reaction diffusion networks by explicitely modeling spatial details of the diffusion steps. That is, in this model we also account for local concentration differences due to reactions (reducing or increasing the number of clusters) or caused by thermal noise. In our simulation we are now able to explore the nature of transitions between the substates. This enables us to gather information about the dynamics of the full model in terms of our network view. Assuming Poisson type dynamics for each transition, the network could then be treated like a Master equation [78]. A similar modeling approach has already been successfully applied to the problem of protein conformation dynamics [96, 97]. However, the transport transition dynamics does not necessarily have to be Poisson like and thus one has to carefully check whether the Master equation approach is valid. In general, the required information about the transition dynamics is hardly accessible in experiment, because it requires knowledge about the position of any of the proteins in a certain region of a sample. Although fluorescence resonance energy transfer microscopy (FRET) [147] allows for the *in vivo* identification of protein-protein interactions, typically only a small fraction of proteins are labeled. As a consequence a specific labeled protein will have many unnoticed interactions. That is, FRET microscopy cannot resolve the dynamics of protein interactions to the desired extent and modeling can be used to fill this gap.

In our model, each partial cluster or encounter complex can be uniquely assigned to one of the substates shown in Fig. 5.6b. At some point it will undergo a transition to one or more different substates. For example, encounter complex $\mathcal{C}4$ could degrade into a separate particle $\mathcal{C}1$ and a two-particle cluster $\mathcal{C}3$ by diffusion. The lifetime of the previous substate is equivalent with the transition time from this substate to the next ones. The transition times T and relative frequencies f are the two quantities which are accessible in the simulation. However, in general there can be more than one possible transition from a particular substate. Therefore, a statement about the underlying kinetics of the transitions cannot be calculated from the transition times straightforwardly. First, let us suppose all tran-

sitions could be modeled by a Poisson process. Let some state $\mathcal{A}$ stochastically switch over to a number N of states $\mathcal{B}_i$ at independent kinetic rates k_i, respectively. What would be the expected transition times T_i and frequencies f_i in the whole system? Starting from state $\mathcal{A}$, the dynamic equations are:

$$\dot{\mathcal{A}} = -\sum_{i=1}^{N} k_i \mathcal{A} = -\mathcal{A}\left(\sum_{i=1}^{N} k_i\right) , \qquad \dot{\mathcal{B}}_i = k_i \mathcal{A} . \tag{5.17}$$

The solution of the dynamic equation for $\mathcal{A}$ is a single exponential:

$$\mathcal{A}(t) = \mathcal{A}_0 \exp\left[-\sum_{i=1}^{N} k_i t\right] . \tag{5.18}$$

As $\mathcal{B}_i$ are dynamically only dependent on $\mathcal{A}$, in the combined system all transitions will happen at the same rate $K = \sum k_i$. That is, the mean transition times will all equally be $T_i = T = 1/K$. The relative frequencies f_i are proportional to the prefactors of $\mathcal{B}_i$ in the steady state, i.e., for large times:

$$\mathcal{B}_i(t) = \int_0^t \mathrm{d}t' \, k_i \mathcal{A}(t) = \mathcal{A}_0 \frac{k_i}{K}(1 - e^{-Kt}) \quad \overset{t\to\infty}{\longrightarrow} \quad \mathcal{A}_0 \frac{k_i}{K} . \tag{5.19}$$

The normalization should be $\sum f_i = 1$, which leads to the finding $f_i = k_i/K$. Knowing only T and f_i from the simulation, the independent kinetic rates of the transitions can therefore be calculated via:

$$k_i = K f_i = f_i / T . \tag{5.20}$$

However, this is only the case, if all transitions can be independently described by Poisson processes.

The simulations of the system described in Fig. 5.6a are parameterized in the spirit of the model $\mathcal{M}1$ from chapter 3, i.e., we consider only a spherical excluded volume effect. Furthermore all parameters are set up dimensionless, i.e., dynamic viscosity $\eta_0 = 1$, thermal energy $k_B T_a = 1$, particle radius $R = 1$ and association rate $k_a = 1$. The patch size is $r = 0.3$ and the distance of the centers of two particles in the bound state is set to $d = 2.55$. Thus, there is an overlapping region between the involved encounter patches in the bound state, i.e., if a bond is broken the particles are still in an encounter complex. On the other hand the overlap is small. This is important as the overlap has to be overcome by diffusion to completely dissociate from the encounter state. Large overlaps could thus lead to an artificially high probability of rebinding especially for high association rates k_a. All simulations are performed at high concentration to confine the run times and mostly high particle numbers to prevent finite size effects. Typically, a periodic boundary box with edge length $L = 28$ contains $N = 192$ particles. To get an impression of the meaning of these values in real microscopic dimensions, we want to deduce the dimensionless time unit (TU) in the system, assuming a typical scale for the length unit $1\mathrm{LU} = 1\mathrm{nm}$. As the diffusion coefficient is given

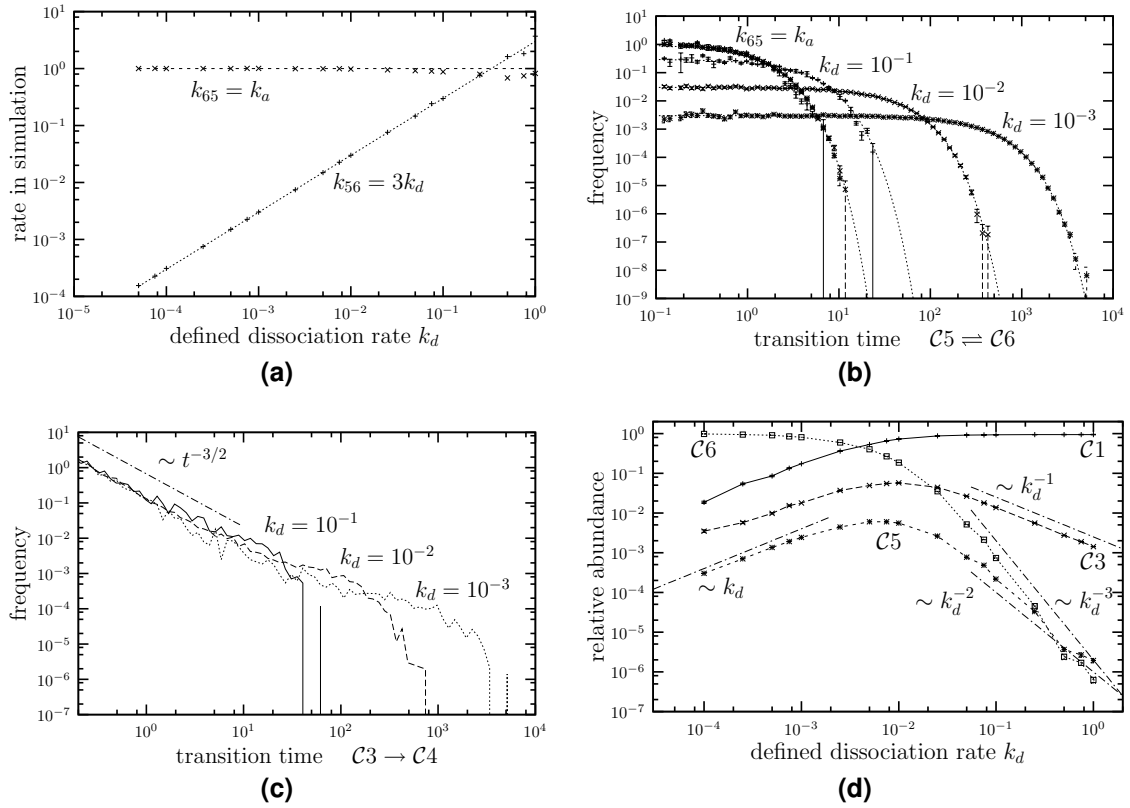


Figure 5.7: (a) Rates between substate $\mathcal{C}5$ and $\mathcal{C}6$ (compare Fig. 5.6b) as obtained from simulations at different predefined dissociation rates k_d represented by the data points plotted along with the expectations in dashed and dotted lines. (b) Transition time distributions between $\mathcal{C}5$ and $\mathcal{C}6$ for three values of the dissociation rate $k_d = \{10^{-1}, 10^{-2}, 10^{-3}\}$. Data points represent the histogram data from the simulations, the dotted lines hint the corresponding ideal Poisson behavior from the expected rates. (c) $\mathcal{C}3 \rightarrow \mathcal{C}4$ transition time distribution for $k_d = 10^{-1}$ (solid line), $k_d = 10^{-2}$ (dashed line), and $k_d = 10^{-3}$ (dotted line). For the sake of clarity the data is shown with lines here. (d) Normalized temporal abundance of the four possible complexes $\mathcal{C}1$, $\mathcal{C}3$, $\mathcal{C}5$, and $\mathcal{C}6$. The data shows the sum of all durations a certain complex exists in the simulation, normalized by the total of the four values.

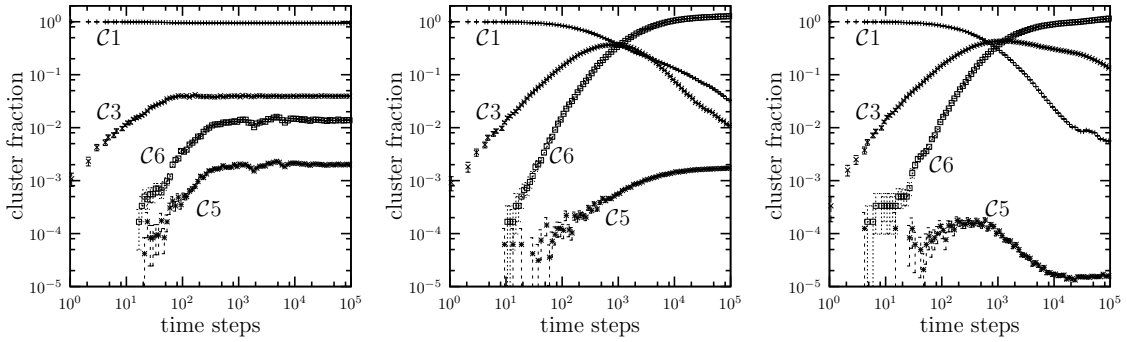
by $D = k_B T_a/(6\pi\eta_0 R)$ it is thus in system units $D = 1/(6\pi)\mathrm{LU}^2/\mathrm{TU}$. Combining these two equations and using real values for the involved constants k_B, $T_a = 293\mathrm{K}$, $\eta_0 = 10^{-3}\mathrm{Pa\,s}$ and $R = 1\mathrm{LU} = 1\mathrm{nm}$ we find $1\mathrm{TU} = (1\mathrm{LU})^2 R\eta_0/k_B T_a \approx 0.25\mathrm{ns}$. Obviously, the unit time in our simulations is in a reasonable regime when the other parameters are scaled like under molecular conditions. Furthermore, with $L = 28\mathrm{nm}$ and $N = 192$ we can estimate the concentration of the system $c = N/(N_A L^3) \approx 14\mathrm{mM}$. This is not unphysical although much higher than the concentrations used in the studies in chapter 3 and 4. However, the assumed reactive association rate of $k_a = 1\mathrm{TU}^{-1} = 4 \cdot 10^9\mathrm{s}^{-1}$ is of course rather large and basically means that the system will behave like diffusion limited.

During simulations, we count the number of certain transitions and average the corresponding transition times. The run is stopped after a total of 10^6 tran-

sitions have occurred. In this regard, the results represent equilibrium rather than first passage properties of the system. All bonds are assumed to form from and break into an encounter complex with the same stochastic rates k_a and k_d, respectively. The equilibrium simulations are performed over a wide range of dissociation rates for the bonds $k_d = 5 \cdot 10^{-5} \dots 1$. Some of the possible transitions shown in Fig. 5.6b exclusively depend on k_a and k_d. This is especially the case for the three-particle cluster $\mathcal{C}5$ and the fully bound cluster $\mathcal{C}6$, which are formed by reaction processes. Particularly, the transition from $\mathcal{C}5$ to $\mathcal{C}6$ requires the formation of one bond ($k_{65} = k_a$), while the opposite direction can happen by the dissociation of one of three bonds ($k_{56} = 3k_d$). In Figs. 5.7a and 5.7b the mean transition rates as obtained by the treatment explained in Eq. (5.20) and some transition time histograms are shown for $\mathcal{C}5 \rightleftharpoons \mathcal{C}6$. The expectations of k_{65} and k_{56} agree with the simulation results. Also, the histogram data corresponds well to the assumed Poisson-like behavior. Particularly, the different transition histogram data of $\mathcal{C}5 \rightarrow \mathcal{C}6$ clearly agree which is reasonable as this transition only depends on the constant association rate $k_a = 1$.

However, we find that as soon as transport processes are involved in the dynamics of a substate, the initial assumption of purely Poisson-type transitions in the network is not fulfilled anymore. This also leads to a break-down of the treatment described in Eq. (5.20). One example is shown in Fig. 5.7c. It shows transition time histograms of the two-particle cluster $\mathcal{C}3$ to an encounter complex with a single particle $\mathcal{C}4$, which occurs due to diffusion. The histograms are not completely exponential but show a power-law behavior $t^{-3/2}$ at small times. This is similar to what we already found for the contact resting and return times in chapter 3, compare Figs. 3.9a and 3.9b. Again, the reasonable explanation is that if the encounter $\mathcal{C}4$ did not successfully lead to a bond formation ($\rightarrow \mathcal{C}5$), the encountering particles do not separate to a far distance, but quickly reform the encounter. This is completely analog to the considerations about the number of contacts before encounter in chapter 3. Actually, the data should be independent of the chosen dissociation rate k_d, as the process is pure transport. However, due to the decay of the two-particle cluster at rate k_d, the abundance times are exponentially decaying, which in turn gives an exponential decay of the encounter formation times shown in the plot. Therefore, it makes sense that the tail of the histograms decrease slower for smaller dissociation rates.

Fig. 5.7d compares the relative temporal abundances of the substates except encounter complexes (i.e. $\mathcal{C}1$, $\mathcal{C}3$, $\mathcal{C}5$, and $\mathcal{C}6$) for different dissociation rates k_d. The data was obtained the total number of times a certain by summing up all durations a certain cluster existed. For each k_d the four obtained values were then normalized by the sum of all duration times. The stoichiometry of the complexes was not considered, i.e., each single particle $\mathcal{C}1$ was counted individually, although on the other hand the maximum number of fully bound complexes $\mathcal{C}6$ was a factor 3 smaller. As we present the data in a double-logarithmic plot, a different counting rule would only shift the data a little in y-direction but would not alter the qualitative behavior. The most obvious finding is that low dissociation rates do kinetically favor the fully bound state $\mathcal{C}6$, while high dissociation rates lead to quick breaking of formed bonds and lead to a high fraction of single par-



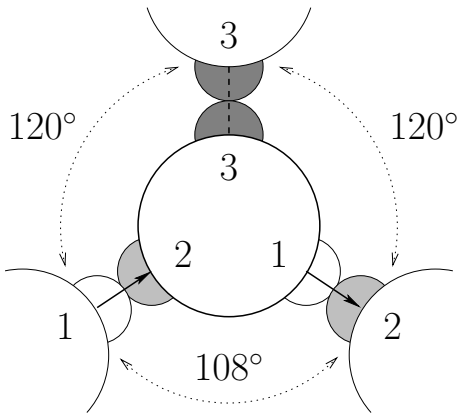
(a) $k_d = 10^{-1}$ **(b)** $k_d = 10^{-3}$ **(c)** $k_d = 10^{-5}$

Figure 5.8: Time series of the fractions of clusters $\mathcal{C}1$, $\mathcal{C}3$, $\mathcal{C}5$, and $\mathcal{C}6$ weighted by their size. The weighting means that if all particles are in $\mathcal{C}6$ clusters, the fraction of $\mathcal{C}6$ would be 1. The three plots show the results for three different dissociation rates: (a) $k_d = 10^{-1}$, (b) $k_d = 10^{-3}$, (c) $k_d = 10^{-5}$.

ticles $\mathcal{C}1$. The abundance of the intermediate complexes $\mathcal{C}3$ and $\mathcal{C}5$ are highest in an intermediate regime, where both $\mathcal{C}1$ and $\mathcal{C}6$ have a similar probability of occurrence.

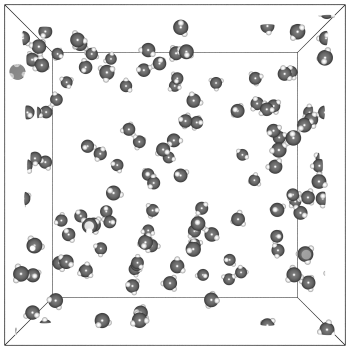
In the unstable regime, i.e., for high $k_d \approx k_a$ the abundance of complexes with n bonds formed decrease with k_d^{-n}. As the stability of the bonds decreases with larger k_d, the probability of finding a two-particle cluster $\mathcal{C}3$ must decrease linearly $\sim k_d^{-1}$. The argument can be extended to substates with $n > 1$ bonds, which are destabilized proportional to k_d^{-1} with respect to some cluster with $n - 1$ bonds. For small k_d, i.e., in the complex stabilizing regime the abundance of $\mathcal{C}6$ saturates while all others decrease approximately proportional to k_d at least in the observed parameter range. In analogy to large k_d, a relative preference of clusters with n bonds proportional to $k_d^{-(3-n)}$ could be expected. However, the transport part of the association process does systematically "stabilze" smaller clusters in the unbound state because there is a certain lag time until they find a suitable binding partner which is only dependent on concentration but not on the reaction rates k_a and k_d.

Finally, in Fig. 5.8 three time series of cluster fractions are shown for different dissociation rates k_d. The data is averaged over 1000 independent simulation runs with 24 model particles which are randomly distributed initially. Therefore, each plot starts with the fraction of $\mathcal{C}1$ $f(\mathcal{C}1) = 1$ and no other substates populated. As clusters with one or more bonds form, the fraction of $\mathcal{C}1$ is reduced. As in the previous consideration only $\mathcal{C}1$, $\mathcal{C}3$, $\mathcal{C}5$, and $\mathcal{C}6$ are counted for the normalization. At some point the system must reach a steady state, where all fractions stay constant on average. In the observed time frame this is only the case for a large dissociation rate $k_d = 10^{-1}$. In contrast, the data shown in Fig. 5.7 was obtained from much longer simulations, where the system was always in a quasi-equilibrium. Here, for $k_d = 10^{-1}$ the fraction of bound clusters is quite low for all times. Thus, the steady state is close to the initial state and is reached quickly. Remarkably, the transport part of the assembly process is clearly changing the

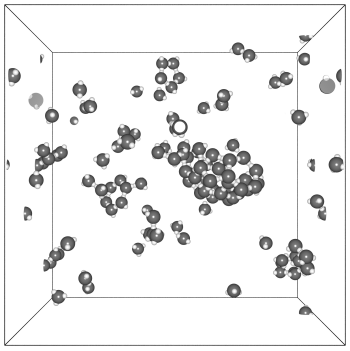


(a) Local rules T1 virus

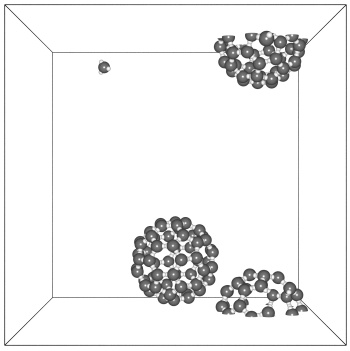
Figure 5.9: (a) Set of local rules for the formation of a T1 virus capsid out of 60 particles of the same type (adopted from Ref. [129]). Each particle has three interaction sites with enclosed angles as given in the figure. As the sum of the angles is smaller than $180°$ it is obvious that the three sites do not lie in one plane with the center of the particle. This leads to the formation of a three-dimensional final structure. The different sites are numbered and colored in different gray scales. Site 1 of one particle can only bind with site 2 of another one, as well as 3 can only bind to 3. (b)–(d) Snapshots from a simulation with 120 particles in our model framework $\mathcal{M}1$ (compare chapter 3). Time is given in simulation steps.



(b) $T = 1000$



(c) $T = 180000$



(d) $T = 3711000$

behavior compared to a system where bonds form and dissociate purely according to k_a and k_d. In the latter, as soon as $k_a > k_d$, the state with the maximum number of bonds would always be the most populated one in the steady state. This is not the case here. However, a certain small amount of fully bound clusters $\mathcal{C}6$ emerges. This leads in turn to a fraction of three-particle clusters with one broken bond $\mathcal{C}5$ which is approximately a factor $k_d/k_a = 10^{-1}$ smaller. This ratio of $f(\mathcal{C}5)/f(\mathcal{C}6) = k_d/k_a$ is found in all three cases. The dynamics in the beginning as well as the time at which the fraction of $\mathcal{C}1$ gets smaller than the fraction of $\mathcal{C}6$ for small k_d is always rather similar. This is due to two facts. Firstly, the spatial effects caused by initial configuration and the concentration of the system are equal in all cases. Secondly, it can be shown that the principal mode of the reaction dynamics if transport is completely neglected is $k_a + k_d$, which is roughly equal in all cases.

5.3 Modeling large complexes – Virus capsids

Next we extend our considerations to large clusters. A class of systems which can be well described by model approaches like ours are virus capsids. In the

introduction a number of models have been reviewed that have been developed to study self-assembly of virus like structures (see Sect. 1.4). All of these studies use Lennard-Jones-type (LJ) potentials in order to steer the system towards bound clusters. Typically, the final structure is predetermined analogously to the local rules theory of Berger et al. [23] by the form of the potential, e.g., by specific orientation dependent LJ terms or additional harmonic terms. Most of the cited references do either have no direct representation of the altered friction properties of growing clusters or use implicit hydrodynamics by considering independent motion of all monomers and have the bound monomers tied to each other by the applied potential. The use of explicit Newtonian dynamics requires the system to be evolved carefully and in small time steps. In contrast, in our model it is possible to use much larger time steps, particularly in the most generic version $\mathcal{M}1$ (compare chapter 3) where no interaction potentials are implied. The lack of simulation speed is one of the reasons, why interesting questions regarding the assembly pathway are not well understood yet. Although the mechanisms of growth in terms of cluster size of added fragments has been addressed at least to some extend [74, 117, 167], this problem has not been studied thoroughly so far. Particularly, the topology of the partially formed clusters is not considered in most cases except in Ref. [117], where static distributions of bond numbers at fixed cluster sizes are indicated.

The local rules approach can be implemented in our model by locating the reactive patches according to the required geometry as illustrated in Fig. 5.9a. The Figs. 5.9b–5.9d show snapshots of a simulation with 120 particles forming two T1 virus capsids in a simulation of about $4 \cdot 10^6$ time steps. As in all theoretical studies of this type, the initial monomer concentration is chosen rather high to speed up the association process. Particularly, the periodic boundary box is cubic and has an edge length of $L = 40$ while the particles have radii $R = 1$ and patch sizes $r = 0.4$. The shown simulation was performed in roughly three hours on a single CPU. While in many similar approaches the amount of computed data restricts to a few trajectories, our fast simulation approach allows for collecting reasonable statistics.

As an example of the performance of our method we perform a similar study as shown in Fig. 5.8 for the three-particle cluster. Initially, we randomly place 60 monomers in a cubic boundary box with an edge length of $L = 30$. We average over 200 simulations of 10^6 time steps each, which makes a total computation time of roughly one day on a single CPU. The study is similar to what Rapaport presented recently [117]. Indeed, our system size is limited to only one fully formed capsid. That is, it will be fragile to trapping in local minima, e.g., if two partially formed clusters use up all particles but do not fit together. In this sense Rapaport stated that unbinding is important for self-assembly to prevent such trapping. Exactly the same is seen in our results. If the dissociation constant k_d is too large like in Fig. 5.10a all bonds do quickly break again and no larger clusters can form. In the other limit of very low dissociation rates (see Fig. 5.10d) the bond formation starts quickly, but the system is trapped in local minima. Particularly, a number of smaller clusters are formed which are structured in a way that prevents them from binding, i.e., they either do not fit or there is only a very special set of pairs of

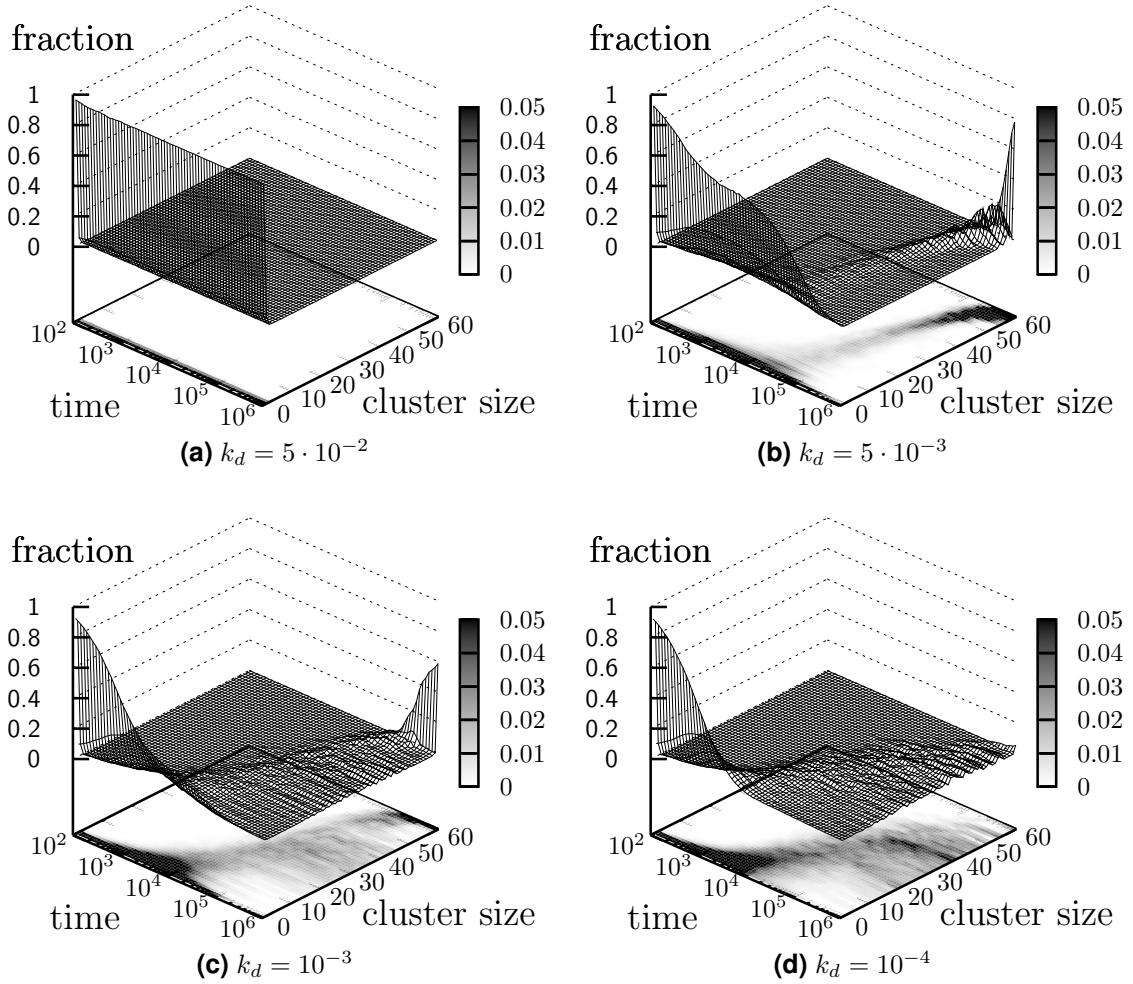


Figure 5.10: Time series of cluster size fractions for the assembly of T1 virus capsid out of 60 monomers. The map at the bottom of the plots is explicitely designed for making the details below a fraction of $f = 0.05$ clearer, which are hard to see in the three dimensional surface representation of the data with lines. The plots show data obtained for different dissociation constants: (a) $k_d = 5 \cdot 10^{-2}$, (b) $k_d = 5 \cdot 10^{-3}$, (c) $k_d = 10^{-3}$, (d) $k_d = 10^{-4}$.

binding sites at which the partially formed clusters could successfully bind. Consequently, while formation of smaller clusters starts at an early stage the cluster size with the highest population after 10^6 simulation steps is at approximately 35 particles. Also a broad range of cluster sizes is roughly equally populated at the end of the runs. In the intermediate range of k_d (see Figs. 5.10c and 5.10b) the formation of larger complexes is delayed to some extend, however, there is a clear preference in the final formation of full clusters with 60 particles. This is especially the case for $k_d = 5 \cdot 10^{-3}$, where there is even more unbinding – and thus less trapping – than for $k_d = 10^{-3}$. Therefore, $k_d = 5 \cdot 10^{-3}$ obviously leads to the most effective assembly scenario of the four observed ones and unbinding at a considerable rate plays an important role.

5.4 Discussion

In this chapter the association dynamics of more than two model proteins has been considered. From the results of three particles on a line it is apparent that the dynamics of subsequent encounter cannot necessarily be separated into two similar transport processes. Moreover, we see in the three dimensional system treated in Sect. 5.2 that it is invalid to describe the transport processes by simple stochastic rates in a network view of assembly. Due to quick re-encounter the transition time histograms obey a power-law at small encounter times. This finding corresponds to what we saw for the contact problem of the bimolecular encounter in chapter 3. Both facts are important because they invalidate the description of assembly with a Master equation approach in a network of transport and reaction processes. Furthermore, the dynamics due to transport and reactions is connected to each other in a way which cannot be understood straightforwardly. The average population of clusters of a certain size depends on the rate constants for binding and unbinding. On the other hand, different concentrations of partially formed clusters will lead to different encounter times, which in turn will affect the amount of higher order clusters. If a single tuneable reaction rate in the network shown in Fig. 5.6b is altered, eventually all transport kinetics are affected. This is basically an additional reason to reject the network approach, at least in the proposed form, as the links are not independent of each other. In summary we can say, that the behavior for, e.g., different parameter sets always has to be analyzed in a full treatment of the model. However, the developed simulation approach is promising because of its specificity as shown in chapter 3 and especially in terms of computation speed, which allows to treat complicated problems. This is exemplarily shown by the study of the assembly of T1 virus capsids at the end of this chapter. We find that capsid formation is most effective for an intermediate range of the dissociation rate, where unbinding is both slow enough to allow formation of larger complexes and quick enough to prevent the system from being trapped in states that do not allow complete capsid formation.

While the virus capsid simulations are able to reproduce known qualitative results there are still a number of issues which have to be considered. From Figs. 5.9b–5.9d it is obvious that while the concentration in terms of particles per simulation volume is constant, the concentration of *free* monomers is strongly decreasing when large clusters are formed. The distances of the particles are distributed around some average value in the beginning, while for nearly completely formed capsids most particles are at a fixed small distance, i.e., bound to each other, and only some freely moving monomers have a comparably large distance from all others. This effective dilution of the largest part of the simulation box is due to the finite size of the system and leads to a strong artificial slow-down in the assembly kinetics.

To our knowledge, in none of the similar theoretical approaches to virus capsid assembly the topologies of important intermediate states have been analyzed. This is one of the interesting open questions and could be achieved by counting populations of particular configurations during the simulation. However, directly identifying a particular topology is rather complicated. Additionally, the number

of possible topologies can be large. Even if one restricts to the information of how many monomers N and how many bonds N_B are in a cluster, N_B can vary between $N-1$ and up to $M \times N$ if each monomer can bind to a maximum number of M neighbors (in case of our T1 structure we have $M = 3$). I.e., the number of distinguishable topologies would be already in the order of $\overline{N}^2$ if $\overline{N}$ is the number of monomers in the completely assembled cluster.

Even more complex is the identification of important pathways in an overall assembly network like considered for the three-particle cluster. The number of possible encounter complexes increases with the second power of the number of possible substates because any substate can encounter any other one even if they are not able to form a bond. Therefore, a transition network taking into account cluster topologies and all encounter to encounter transitions would span a transition matrix with $\sim ((\overline{N}^{\,2})^2)^2 = N^{16}$ entries. In case of the T1 virus with $\overline{N} = 60$ proteins in the final cluster this is a number in the order of $60^{16} = 2.8 \cdot 10^{28}$ which exceeds any available computational resources. Thus, it is necessary to develop more efficient procedures, e.g., by computing and storing only data which is relevant for answering specific questions.

Another interesting subject that can be investigated with our method is the self-assembly dynamics of the Arp 2/3 complex (see Fig. 1.1c). The atomistic structure of Arp 2/3 is known from the work by Robinson et al. [119]. Therefore, the general methods to evaluate suitable parameters for our model can be applied straightforwardly to the structure data. Our description might help to understand, e.g., whether there are particular subclusters that have to form first, before they finally combine into the full complex.

Bibliography

[1] ABRAMOWITZ, M., AND STEGUN, I. *Handbook of mathematical functions*. Dover books on mathematics. Dover Publ., New York, 1972.

[2] AHMAD, M., GU, W., AND HELMS, V. Mechanism of fast peptide recognition by sh3 domains. *Angewandte Chemie* (2008).

[3] ALAMANOVA, D., HELMS, V., ET AL. Free energy landscapes for protein-protein association computed from continuum models. In preparation.

[4] ALBER, F., DOKUDOVSKAYA, S., VEENHOFF, L., ZHANG, W., KIPPER, J., DEVOS, D., SUPRAPTO, A., KARNI-SCHMIDT, O., WILLIAMS, R., CHAIT, B., ET AL. Determining the architectures of macromolecular assemblies. *Nature 450* (Nov. 2007), 683–694.

[5] ALBER, F., DOKUDOVSKAYA, S., VEENHOFF, L. M., ZHANG, W., KIPPER, J., DEVOS, D., SUPRAPTO, A., KARNI-SCHMIDT, O., WILLIAMS, R., CHAIT, B. T., SALI, A., AND ROUT, M. P. The molecular architecture of the nuclear pore complex. *Nature 450*, 7170 (Nov. 2007), 695–701.

[6] ALBERTS, B., BRAY, D., AND LEWIS, J. *Molecular Biology of the Cell*, 4th ed. Taylor & Francis, New York, 2002.

[7] ALFANO, S., AND GREER, M. Determining If Two Solid Ellipsoids Intersect. *Journal of Guidance, Control, and Dynamics 26*, 1 (2003), 106–110.

[8] ALLEN, M., EVANS, G., FRENKEL, D., AND MULDER, B. Hard convex body fluids. *Adv. Chem. Phys. 86* (1993), 1–166.

[9] ALLEN, M. P. Simulations and phase behaviour of liquid crystals. In *Observation, Prediction and Simulation of Phase Transitions in Complex Fluids* (1995), M. Baus, L. Rull, and J. Ryckaert, Eds., vol. 460 of *NATO Science Series C*, Kluwer Academic Publishers, pp. 557–590.

[10] ALLEN, M. P., FRENKEL, D., AND TALBOT, J. Molecular dynamics simulation using hard particles. *Comput. Phys. Rep. 9* (1989), 301–353.

[11] ALSALLAQ, R., AND ZHOU, H.-X. Energy Landscape and Transition State of Protein-Protein Association. *Biophys. J. 92*, 5 (Mar. 2007), 1486–1502.

[12] ALSALLAQ, R., AND ZHOU, H.-X. Prediction of Protein-Protein Asssociation Rates from a Transition-State Theory. *Structure 15* (Feb. 2007), 215–224.

[13] ANDERSON, E., BAI, Z., BISCHOF, C., BLACKFORD, S., DEMMEL, J., DONGARRA, J., DU CROZ, J., GREENBAUM, A., HAMMARLING, S., MCKENNEY, A., AND SORENSEN, D. *LAPACK Users' Guide*, third ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.

[14] ANDRÉ, I., STRAUSS, C., KAPLAN, D., BRADLEY, P., AND BAKER, D. Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl. Acad. Sci. USA 105*, 42 (2008), 16148.

[15] ANTOSIEWICZ, J., BRIGGS, J., AND MCCAMMON, J. Orientational steering in enzyme-substrate association: Ionic strength dependence of hydrodynamic torque effects. *Eur. Biophys. J. 24*, 3 (1996), 137–141.

[16] ARAGON, S. A precise boundary element method for macromolecular transport properties. *J. Comp. Chem.* (2004), 1191–1205.

[17] ARFKEN, G. B., AND WEBER, H. J. *Mathematical Methods for Physicists*, 6th ed. Elsevier Academic Press, Oxford, 2005.

[18] ARKHIPOV, A., FREDDOLINO, P., AND SCHULTEN, K. Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling. *Structure 14*, 12 (2006), 1767–1777.

[19] BARABASI, A.-L., AND OLTVAI, Z. N. Network biologoy: understanding the cell's functional organization. *Nat. Rev. Genet. 5*, 2 (Feb. 2004), 101–113.

[20] BARZYKIN, A., AND SHUSHIN, A. Effect of Anisotropic Reactivity on the Rate of Diffusion-Controlled Reactions: Comparative Analysis of the Models of Patches and Hemispheres. *Biophys. J. 80*, 5 (2001), 2062–2073.

[21] BELL, G. I. Models for the specific adhesion of cells to cells. *Science 200*, 4342 (1978), 618–627.

[22] BERG, H. C., AND PURCELL, E. M. Physics of chemoreception. *Biophys. J. 20* (1977), 193–219.

[23] BERGER, B., SHOR, P., TUCKER-KELLOGG, L., AND KING, J. Local Rule-Based Theory of Virus Shell Assembly. *Proc. Natl. Acad. Sci. USA 91*, 16 (1994), 7732–7736.

[24] BERMAN, H., HENRICK, K., AND NAKAMURA, H. Announcing the worldwide protein data bank. *Nat. Struct. Mol. Biol. 10*, 12 (Dec. 2003), 980. Link: `http://pdb.rcsb.org/`; Statistics: `http://pdb.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/`.

[25] BERNAL, J. M. G., AND DE LA TORRE, J. G. Transport properties and hydrodynamic centers of rigid macromolecules with arbitrary shapes. *Biopolymers 19* (1980), 751–766.

[26] BICOUT, D., AND FIELD, M. Stochastic dynamics simulations of macromolecular diffusion in a model of the cytoplasm of Escherichia coli. *J. Phys. Chem. 100* (1996), 2489–2497.

[27] BLUM, A. S., SOTO, C. M., WILSON, C. D., COLE, J. D., KIM, M., GNADE, B., CHATTERJI, A., OCHOA, W. F., LIN, T., JOHNSON, J. E., AND RATNA, B. R. Cowpea mosaic virus as a scaffold for 3-d patterning of gold nanoparticles. *Nano Lett. 4*, 5 (May 2004), 867–870.

[28] BOX, G. E. P., AND MULLER, M. E. A note on the generation of random normal deviates. *Ann. Math. Stat. 29*, 2 (1958), 610–611.

[29] BRENNER, H., AND O'NEILL, M. On the Stokes resistance of multiparticle systems in a linear shear field. *Chem. Eng. Sci 27* (1972), 1421–1439.

[30] BUCKLE, A. M., SCHREIBER, G., AND FERSHT, A. R. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-.ang. resolution. *Biochemistry 33*, 30 (1994), 8878–8889.

[31] BYRON, O. Construction of hydrodynamic bead models from high-resolution X-ray crystallographic or nuclear magnetic resonance data. *Biophys. J. 72*, 1 (1997), 408–415.

[32] CARRASCO, B., AND DE LA TORRE, J. G. Hydrodynamic properties of rigid particles: Comparison of different modeling and computational procedures. *Biophys. J. 75* (June 1999), 3044–3057.

[33] CARRASCO, B., AND DE LA TORRE, J. G. Improved hydrodynamic interaction in macromolecular bead models. *J. Chem. Phys. 111*, 10 (Sept. 1999), 4817–4826.

[34] CARSLAW, H. S., AND JAEGER, J. C. *Conduction of heat in solids*, 2nd ed. University Press, Oxford, 1986.

[35] CHELAKKOT, R., LIPOWSKY, R., AND GRUHN, T. Novel low-density structure for hard rods with adhesive end groups. *Macromolecules 39*, 20 (October 2006), 7138–7143.

[36] CHONG, L., DEMPSTER, S., HENDSCH, Z., LEE, L., AND TIDOR, B. Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Prot. Sci. 7*, 1 (1998), 206.

[37] COGGAN, J. S., BARTOL, T. M., ESQUENAZI, E., STILES, J. R., LAMONT, S., MARTONE, M. E., BERG, D. K., ELLISMAN, M. H., AND SEJNOWSKI, T. J. Evidence for ectopic neurotransmission at a neuronal synapse. *Science 309*, 5733 (July 2005), 446–451.

[38] DAVIS, M. The slow translation and rotation of two unequal spheres in a viscous fluid. *Chem. Eng. Sci. 24* (1969), 1769–1776.

[39] DE LA TORRE, J. G., AND BLOOMFIELD, V. A. Hydrodynamic properties of macromolecular complexes. i. translation. *Biopolymers 16* (1977), 1747–1763.

[40] DE LA TORRE, J. G., HUERTAS, M. L., AND CARRASCO, B. Calculating hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J. 78*, 2 (Feb. 2000), 719–730. See: `http://leonardo.fcu.um.es/macromol`.

[41] DE LA TORRE, J. G., AND RODES, V. Effects from bead size and hydrodynamic interactions on the translational and rotational coefficients of macromolecular bead models. *J. Chem. Phys. 79*, 5 (Sept. 1983), 2454–2460.

[42] DE LICHTENBERG, U., JENSEN, L. J., BRUNAK, S., AND BORK, P. Dynamic Complex Formation During the Yeast Cell Cycle. *Science 307*, 5710 (2005), 724–727.

[43] DEBYE, P., AND HÜCKEL, E. Zur theorie der elektrolyte II. Das Grenzgesetz für die elektrische leifähigkeit. *Physikalische Zeitschrift 24*, 15 (1923), 305–325.

[44] DEBYE, P. J. W. Reaction rates in ionic solutions. *Trans. Electrochem. Soc. 82* (1942), 265–272.

[45] DELISI, C., AND WIEGEL, F. Effect of Nonspecific Forces and Finite Receptor Number on Rate Constants of Ligand–Cell Bound-Receptor Interactions. *Proc. Natl. Acad. Sci. USA 78*, 9 (1981), 5569–5572.

[46] DONG, F., VIJAYAKUMAR, M., AND ZHOU, H. Comparison of Calculation and Experiment Implicates Significant Electrostatic Contributions to the Binding Stability of Barnase and Barstar. *Biophys. J. 85*, 1 (2003), 49–60.

[47] EIGEN, M. Diffusion control in biochemical reactions. *Quant. Stat. Mech. Nat. Sci.* (1974), 37–61.

[48] EINSTEIN, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys. 322*, 8 (1905), 549–560.

[49] ELCOCK, A., GABDOULLINE, R., WADE, R., AND MCCAMMON, J. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J. Mol. Bio. 291*, 1 (1999), 149–162.

[50] ELCOCK, A., SEPT, D., AND MCCAMMON, J. Computer Simulation of Protein-Protein Interactions. *J. Phys. Chem. B 105*, 8 (2001), 1504–1518.

[51] ELTIS, L. D., HERBERT, R. G., BAKER, P. D., MAUK, A. G., AND NORTHRUP, S. H. Reduction of Horse Heart Ferricytochrome c by Bovine Liver Ferrocytochrome b_5. Experimental and Theoretical Analysis. *Biochemistry 30*, 15 (1991), 3663–3674.

[52] ERDMANN, T., AND SCHWARZ, U. Impact of receptor-ligand distance on adhesion cluster stability. *Eur. Phys. J. E 22*, 2 (2007), 123–137.

[53] ERMAK, D., AND MCCAMMON, J. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys. 69* (1978), 1352.

[54] FANGE, D., AND ELF, J. Noise-induced min phenotypes in e. coli. *PLoS Comp. Bio. 2*, 6 (June 2006), 0637–0648.

[55] FANTONI, R., GAZZILLO, D., GIACOMETTI, A., MILLER, M. A., AND PASTORE, G. Patchy sticky hard spheres: Analytical study and monte carlo simulations. *J. Chem. Phys. 127*, 23 (Dec. 2007), 234507.

[56] FRANK, J. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys. Biomol. Struct. 31* (2002), 303–319.

[57] GABDOULLINE, R., AND WADE, R. Biomolecular diffusional association. *Curr. Op. Struct. Bio. 12*, 2 (2002), 204–213.

[58] GABDOULLINE, R. R., AND WADE, R. C. Simulation of diffusional association of barnase and barstar. *Biophys. J. 72* (May 1997), 1917–1929.

[59] GAVIN, A., ALOY, P., GRANDI, P., KRAUSE, R., BOESCHE, M., MARZIOCH, M., RAU, C., JENSEN, L., BASTUCK, S., DÜMPELFELD, B., ET AL. Proteome survey reveals modularity of the yeast cell machinery. *Nature 440* (Mar. 2006), 631–636.

[60] GAVIN, A., BÖSCHE, M., KRAUSE, R., GRANDI, P., MARZIOCH, M., BAUER, A., SCHULTZ, J., RICK, J., MICHON, A., CRUCIAT, C., ET AL. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature 415* (Jan. 2002), 141–147.

[61] GEIGER, B., BERSHADSKY, A., PANKOV, R., AND YAMADA, K. M. Transmembrane crosstalk between the extracellular matrix and the cytoskeleton. *Nat. Rev. Mol. Cell Bio. 2*, 11 (Nov. 2001), 793–805.

[62] GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics 22*, 4 (Dec. 1976), 403–434.

[63] GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry 81*, 25 (Dec. 1977), 2340–2361.

[64] GLOTZER, S. C. Materials science: Some assembly required. *Science 306*, 5695 (Oct. 2004), 419–420.

[65] GO, N. Theoretical Studies of Protein Folding. *Annual Reviews in Biophysics and Bioengineering 12*, 1 (1983), 183–210.

[66] GOEL, N. S., AND RICHTER-DYN, N. *Stochastic Models in Biology*, 1st ed. Academic Press, New York, 1974.

[67] GOLDMAN, A., COX, R., AND BRENNER, H. The slow motion of two identical arbitrarily oriented spheres through a viscous fluid. *Chem. Eng. Sci. 21* (1966), 1151–1170.

[68] GOLDSTEIN, B., COOMBS, D., HE, X., PINEDA, A., AND WOFSY, C. The influence of transport on the kinetics of binding to surface receptors: application to cells and BIAcore. *J. Mol. Recognit 12* (1999), 293–299.

[69] GOLDSTEIN, B., AND DEMBO, M. Approximating the effects of diffusion on reversible reactions at the cell surface: ligand-receptor kinetics. *Biophys. J. 68*, 4 (1995), 1222–1230.

[70] GOODSELL, D. Inside a living cell. *Trends in Biochemical Sciences 16* (1991), 203–206.

[71] GORBA, C., GEYER, T., AND HELMS, V. Brownian dynamics simulations of simplified cytochrome c molecules in the presence of a charged surface. *J. Chem. Phys. 121*, 1 (July 2004), 457–464.

[72] GRACIAS, D. H., BONCHEVA, M., OMOREGIE, O., AND WHITESIDES, G. M. Biomimetic self-assembly of helical electrical circuits using orthogonal capillary interactions. *Appl. Phys. Lett. 80*, 15 (2002), 2802–2804.

[73] HAGAN, M. Controlling viral capsid assembly with templating. *Physical Review E 77*, 5 (2008), 51904.

[74] HAGAN, M., AND CHANDLER, D. Dynamic Pathways for Viral Capsid Assembly. *Biophys. J. 91*, 1 (2006), 42.

[75] HAN, Y., ALSAYED, A. M., NOBILI, M., ZHANG, J., LUBENSKY, T. C., AND YODH, A. G. Brownian motion of an ellipsoid. *Science 314* (Oct. 2006), 626–630.

[76] HARVEY, S. C., AND DE LA TORRE, J. G. Coordinate systems for modeling the hydrodynamic resistance and diffusion coefficients of irregularly shaped rigid macromolecules. *Macromolecules 13*, 4 (July 1980), 960–964.

[77] HELMS, V. *Principles of Computational Cell Biology: From Protein Complexes to Cellular Networks*, 1st ed. Wiley-Vch, June 2008.

[78] HONERKAMP, J. *Stochastische Dynamische Systeme*, 1st ed. VCH, Weinheim, 1990.

[79] IWAHARA, J., AND CLORE, G. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature 440*, 7088 (2006), 1227–1230.

[80] KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., WYCKOFF, H., AND PHILLIPS, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature 181*, 4610 (Mar. 1958), 662–666.

[81] KENTSIS, A., AND BORDEN, K. L. B. Physical mechanisms and biological significance of supramolecular protein self-assembly. *Curr. Protein Pept. Sci. 5* (2004), 125–134.

[82] KIM, Y., TANG, C., CLORE, G., AND HUMMER, G. Replica exchange simulations of transient encounter complexes in protein–protein association. *Proc. Natl. Acad. Sci. USA 105*, 35 (2008), 12855–12860.

[83] KIRKWOOD, J. G. The statistical mechanical theory of irreversible processes in solutions of flexible macromolecules, visco-elastic behavior. *Rec. Trav. Chim. 68* (1949), 649–660.

[84] KIRKWOOD, J. G., AND RISEMAN, J. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys. 16*, 6 (June 1948), 565–573.

[85] KORN, C. B., AND SCHWARZ, U. S. Efficiency of initiating cell adhesion in hydrodynamic flow. *Phys. Rev. Lett. 97* (Sept. 2006), 138103–1–4.

[86] KORN, C. B., AND SCHWARZ, U. S. Mean first passage times for bond formation for a brownian particle in linear shear flow above a wall. *J. Chem. Phys. 126* (Mar. 2007), 095103–1–15.

[87] KORN, C. B., AND SCHWARZ, U. S. Mean encounter times for cell adhesion in hydrodynamic flow: Analytical progress by dimensional reduction. *Eur. Phys. Lett. 83*, 2 (2008), 28007.

[88] KUSSIE, P. H., GORINA, S., MARECHAL, V., ELENBAAS, B., MOREAU, J., LEVINE, A. J., AND PAVLETICH, N. P. Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science 274*, 5289 (1996), 948–953.

[89] LAGERHOLM, B., AND THOMPSON, N. Temporal Dependence of Ligand Dissociation and Rebinding at Planar Surfaces. *J. Phys. Chem. B 104*, 4 (2000), 863–868.

[90] LANGE, O. F., LAKOMEK, N.-A., FARES, C., SCHRODER, G. F., WALTER, K. F. A., BECKER, S., MEILER, J., GRUBMULLER, H., GRIESINGER, C., AND DE GROOT, B. L. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science 320*, 5882 (2008), 1471–1475.

[91] LANGEVIN, P. The recombination and mobilities of ions in gases. In *Annales de Chimie et de Physique* (1903), vol. 28, pp. 433–530.

[92] LAUFFENBURGER, D., AND LINDERMAN, J. *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford University Press, USA, 1993.

[93] MCGUFFEE, S., AND ELCOCK, A. Atomically detailed simulations of concentrated protein solutions: The effects of salt, ph, point mutations, and protein concentration in simulations of 1000-molecule systems. *J. Am. Chem. Soc. 128*, 37 (2006), 12098–12110.

[94] MITRAGOTRI, S., AND LAHANN, J. Physical approaches to biomaterial design. *Nat. Mater. 8*, 1 (Jan. 2009), 15–23.

[95] MIYASHITA, O., ONUCHIC, J., AND OKAMURA, M. Transition state and encounter complex for fast association of cytochrome c 2 with bacterial reaction center. *Proc. Natl. Acad. Sci. USA 101*, 46 (2004), 16174–16179.

[96] NOÉ, F., HORENKO, I., SCHÜTTE, C., AND SMITH, J. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys. 126* (2007), 155102–1–17.

[97] NOE, F., KRACHTUS, D., SMITH, J., AND FISCHER, S. Transition networks for the comprehensive characterization of complex conformational change in proteins. *J. Chem. Theory and Comput 2* (2006), 840–857.

[98] NORTHRUP, S. Diffusion-controlled ligand binding to multiple competing cell-bound receptors. *J. Phys. Chem. 92*, 20 (1988), 5847–5850.

[99] NORTHRUP, S., BOLES, J., AND REYNOLDS, J. Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science 241*, 4861 (1988), 67–70.

[100] NORTHRUP, S., AND ERICKSON, H. Kinetics of Protein-Protein Association Explained by Brownian Dynamics Computer Simulation. *Proc. Natl. Acad. Sci. USA 89*, 8 (1992), 3338–3342.

[101] NORTHRUP, S. H., ALLISON, S. A., AND MCCAMMON, J. A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys. 80*, 4 (1984), 1517–1524.

[102] OLSON, D. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct 29* (2000), 105–153.

[103] ONSAGER, L. Zur Theorie der Elektrolyte I. *Phys. Z. 27* (1926), 388.

[104] ONSAGER, L. Zur Theorie der Elektrolyte II. *Phys. Z. 28* (1927), 277.

[105] PAINTER, A., JAYA, N., BASHA, E., VIERLING, E., ROBINSON, C., AND BENESCH, J. Real-Time Monitoring of Protein Complexes Reveals their Quaternary Organization and Dynamics. *Chem. Bio. 15*, 3 (2008), 246–253.

[106] PATTERSON, S. D., AND AEBERSOLD, R. H. Proteomics: the first decade and beyond. *Nature Genetics 33* (2003), 311–323.

[107] PELLETIER, H., AND KRAUT, J. Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c. *Science 258*, 5089 (1992), 1748–1755.

[108] PERRAM, J., AND WERTHEIM, M. Statistical mechanics of hard ellipsoids. I: Overlap algorithm and the contact function. *J. Comput. Phys. 58*, 3 (1985), 409–416.

[109] PERRAM, J., WERTHEIM, M., LEBOWITZ, J., AND WILLIAMS, G. Monte Carlo simulation of hard spheroids. *Chem. Phys. Lett. 105*, 3 (1984), 277–280.

[110] PERRIN, F. Mouvement brownien d'un ellipsoïde (I): Dispersion diélectrique pour des molécules ellipsoïdales. *J. Phys. Radium 5* (1934), 497–511.

[111] PERRIN, F. Mouvement brownien d'un ellipsoïde (II): Rotation libre et depolarisation des flourescences. Translation et diffusion des molécules ellipsoïdales. *J. Phys. Radium 7* (1936), 1–11.

[112] PÓLYA, G. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Math. Ann. 84*, 1–2 (Mar. 1921), 149–160.

[113] POLYANIN, A. *Handbook of Linear Partial Differential Equations for Engineers and Scientists*. Russian Academy of Sciences, Moscow, 2001.

[114] RAI, N., NÖLLMANN, M., SPOTORNO, B., TASSARA, G., BYRON, O., AND ROCCO, M. SOMO (SOlution MOdeler): Differences between X-Ray- and NMR-Derived Bead Models Suggest a Role for Side Chain Flexibility in Protein Hydrodynamics. *Structure 13* (May 2005), 723–734.

[115] RAMANATHAN, S., AND MORSE, D. C. Brownian dynamics algorithm for entangled wormlike threads. *J. Chem. Phys. 126* (Mar. 2007), 094906–1–11.

[116] RAPAPORT, D. C. Self-assembly of polyhedral shells: A molecular dynamics study. *Phys. Rev. E 70*, 5 (2004), 51905.

[117] RAPAPORT, D. C. Role of reversibility in viral capsid growth: A paradigm for self-assembly. *Phys. Rev. Lett. 101*, 18 (2008), 18610–1–4.

[118] REDNER, S. *A Guide to First-Passage Processes*, 1st ed. University Press, Cambridge, 2001.

[119] ROBINSON, R. C., TURBEDSKY, K., KAISER, D. A., MARCHAND, J. B., HIGGS, H. N., CHOE, S., AND POLLARD, T. D. Crystal structure of arp2/3 complex. *Science 294*, 5547 (Nov. 2001), 1679–1684.

[120] ROTH, R., VAN ROIJ, R., ANDRIENKO, D., MECKE, K. R., AND DIETRICH, S. Entropic torque. *Phys. Rev. Lett. 89*, 8 (July 2002), 088301–1–4.

[121] ROTNE, J., AND PRAGER, S. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys. 50*, 11 (June 1969), 4831–4837.

[122] SALI, A., GLAESER, R., EARNEST, T., AND BAUMEISTER, W. From words to literature in structural proteomics. *Nature 422* (Mar. 2003), 216–225.

[123] SCHLOSSHAUER, M., AND BAKER, D. A general expression for bimolecular association rates with orientational constraints. *J. Phys. Chem. B 106*, 46 (2002), 12079–12083.

[124] SCHLUTTIG, J., ALAMANOVA, D., HELMS, V., AND SCHWARZ, U. S. Dynamics of protein-protein encounter: A Langevin equation approach with reaction patches. *J. Chem. Phys. 129* (2008), 155106.

[125] SCHREIBER, G., AND FERSHT, A. R. Interaction of Barnase with its Polypeptide Inhibitor Barstar Studied by Protein Engineering. *Biochemistry 32*, 19 (1993), 5145–5150.

[126] SCHREIBER, G., AND FERSHT, A. R. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Bio. 3*, 5 (May 1996), 427–431.

[127] SCHREIBER, G., HARAN, G., AND ZHOU, H. X. Fundamental aspects of protein-protein association kinetics. *Chem. Rev. 0*, 0 (2009).

[128] SCHWARTZ, R., GARCEA, R., AND BERGER, B. "Local Rules" Theory Applied to Polyomavirus Polymorphic Capsid Assemblies. *Virology 268*, 2 (2000), 461–470.

[129] SCHWARTZ, R., SHOR, P., PREVELIGE, P., AND BERGER, B. Local Rules Simulation of the Kinetics of Virus Capsid Self-Assembly. *Biophys. J. 75*, 6 (1998), 2626–2636.

[130] SCIORTINO, F. Disordered materials: One liquid, two glasses. *Nat. Mater. 1*, 3 (Nov. 2002), 145–146.

[131] SELZER, T., AND SCHREIBER, G. Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. *J. Mol. Bio. 287*, 2 (1999), 409–419.

[132] SHARON, M., AND ROBINSON, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Ann. Rev. Biochem. 76* (July 2007), 167–193.

[133] SHEINERMAN, F., AND HONIG, B. On the Role of Electrostatic Interactions in the Design of Protein–Protein Interfaces. *J. Mol. Bio. 318*, 1 (2002), 161–177.

[134] SHIMIZU, H. Effect of molecular shape on nuclear magnetic resonance. *J. Chem. Phys. 37*, 4 (Aug. 1962), 765–778.

[135] SHOUP, D., LIPARI, G., AND SZABO, A. Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys. J. 36*, 3 (1981), 697–714.

[136] SHOUP, D., AND SZABO, A. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J. 40*, 1 (1982), 33–39.

[137] SHUSHIN, A., AND BARZYKIN, A. Effect of Local Molecular Shape and Anisotropic Reactivity on the Rate of Diffusion-Controlled Reactions. *Biophys. J. 81*, 6 (2001), 3137–3145.

[138] SMOLUCHOWSKI, M. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys. 326*, 14 (1906), 756–780.

[139] SMOLUCHOWSKI, M. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. Phys. Chem. 92* (1917), 129–168.

[140] SPAAR, A., DAMMER, C., GABDOULLINE, R. R., WADE, R. C., AND HELMS, V. Diffusional encounter of barnase and barstar. *Biophys. J. 90*, 6 (Mar. 2006), 1913–1924.

[141] SPAAR, A., AND HELMS, V. Free energy landscape of protein-protein encounter resulting from brownian dynamics simulations of barnase:barstar. *J. Chem. Theory Comput. 1*, 4 (May 2005), 723–736.

[142] SPECK, T., AND SEIFERT, U. Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state. *EPL (Europhysics Letters) 74*, 3 (2006), 391–396.

[143] STIMSON, M., AND JEFFERY, G. The Motion of Two Spheres in a Viscous Fluid. *Proc. R. Soc. London Ser. A 111*, 757 (1926), 110–116.

[144] SZABO, A., SCHULTEN, K., AND SCHULTEN, Z. First passage time approach to diffusion controlled reactions. *J. Chem. Phys. 72*, 8 (1980), 4350–4357.

[145] TAKADA, S. Gō-ing for the prediction of protein folding mechanisms. *Proc. Natl. Acad. Sci. USA 96*, 21 (1999), 11698.

[146] TOMITA, M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotech. 19*, 6 (June 2001), 205–210.

[147] TRUONG, K., AND IKURA, M. The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Curr. Op. Struct. Bio. 11*, 5 (2001), 573–578.

[148] VAN KAMPEN, N. G. *Stochastic Processes in Physics and Chemistry*. Elsevier, 1992.

[149] VAN ZON, J. S., AND TEN WOLDE, P. R. Green's function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys. 123* (Dec. 2005), 234910–1–16.

[150] VAN ZON, J. S., AND TEN WOLDE, P. R. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett. 94* (Apr. 2005), 128103–1–4.

[151] VENTER, J. C. The sequence of the human genome. *Science 291* (2001), 1304–1351.

[152] VIEILLARD-BARON, J. Phase transitions of the classical hard-ellipse system. *J. Chem. Phys. 56*, 10 (1972), 4729–4744.

[153] VIEILLARD-BARON, J. The equation of state of a system of hard spherocylinders. *Molec. Phys. 28*, 3 (1974), 809–818.

[154] VOET, D., AND VOET, J. *Biochemistry*, 3rd ed. John Wiley & Sons, Hoboken, NJ, 2004.

[155] WADE, R., GABDOULLINE, R., LÜDEMANN, S., AND LOUNNAS, V. Colloquium Paper Electrostatic steering and ionic tethering in enzyme–ligand binding: Insights from simulations. *Proc. Natl. Acad. Sci. USA 95*, 11 (1998), 5942–5949.

[156] WAKIYA, S. Slow motions of a viscous fluid around two spheres. *J. Phys. Soc. Japan 22* (1967), 1101–1109.

[157] WANG, T., TOMIC, S., GABDOULLINE, R., AND WADE, R. How Optimal Are the Binding Energetics of Barnase and Barstar? *Biophys. J. 87*, 3 (2004), 1618–1630.

[158] WANG, W., WANG, J., AND KIM, M.-S. An algebraic condition for the separation of two ellipsoids. *Computer Aided Geometric Design 18* (2001), 531–539.

[159] WILBER, A., DOYE, J., LOUIS, A., NOYA, E., MILLER, M., AND WONG, P. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J. Chem. Phys. 127* (2007), 085106.

[160] WOFSY, C., AND GOLDSTEIN, B. Effective Rate Models for Receptors Distributed in a Layer above a Surface: Application to Cells and Biacore. *Biophys. J. 82*, 4 (2002), 1743–1755.

[161] WÜTHRICH, K. Protein structure determination in solution by nmr spectroscopy. *J. Bio. Chem. 265*, 36 (Dec. 1990), 22059–22062.

[162] WYNNE, S. The crystal structure of the human hepatitis b virus capsid. *Molecular Cell 3*, 6 (June 1999), 771–780. Image from: `http://www.cgl.ucsf.edu/Research/virus/capsids/viruses.html`.

[163] YAMAKAWA, H. Transport properties of polymer chains in dilute solution: Hydrodynamic interaction. *J. Chem. Phys. 53*, 1 (July 1970), 436–443.

[164] YUSUPOV, M. M., YUSUPOVA, G. Z., BAUCOM, A., LIEBERMAN, K., EARNEST, T. N., CATE, J. H. D., AND NOLLER, H. F. Crystal structure of the ribosome at 5.5 a resolution. *Science 292*, 5518 (May 2001), 883–896. Image from: `http://rna.ucsc.edu/rnacenter/ribosome_images.html`.

[165] ZANDI, R., REGUERA, D., BRUINSMA, R., GELBART, W., AND RUDNICK, J. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA 101*, 44 (2004), 15556–15560.

[166] ZANDI, R., VAN DER SCHOOT, P., REGUERA, D., KEGEL, W., AND REISS, H. Classical Nucleation Theory of Virus Capsids. *Biophys. J. 90*, 6 (2006), 1939–1948.

[167] ZHANG, T., AND SCHWARTZ, R. Simulation Study of the Contribution of Oligomer/Oligomer Binding to Capsid Assembly Kinetics. *Biophys. J. 90*, 1 (2006), 57–64.

[168] ZHANG, Y. Progress and challenges in protein structure prediction. *Curr. Op. Struct. Bio. 18*, 3 (June 2008), 342–348.

[169] ZHANG, Z., AND GLOTZER, S. C. Self-assembly of patchy particles. *Nano Lett. 4*, 8 (Aug. 2004), 1407–1413.

[170] ZHOU, H. X. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J. 64*, 6 (1993), 1711–1726.

[171] ZHOU, H. X. Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. *Biophys. J. 73*, 5 (1997), 2441–2445.

[172] ZHOU, H. X., AND SZABO, A. Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions. *Biophys. J. 71*, 5 (1996), 2440–2457.

[173] ZLOTNICK, A. Theoretical aspects of virus capsid assembly. *J. Mol. Rec. 18*, 6 (2005), 479–490.

[174] ZLOTNICK, A., CERES, P., SINGH, S., AND JOHNSON, J. A Small Molecule Inhibits and Misdirects Assembly of Hepatitis B Virus Capsids. *J. Virol. 76*, 10 (2002), 4848–4854.

[175] ZLOTNICK, A., JOHNSON, J., WINGFIELD, P., STAHL, S., AND ENDRES, D. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry 38*, 44 (1999), 14644–14652.

[176] ZWANZIG, R. Diffusion-Controlled Ligand Binding to Spheres Partially Covered by Receptors: An Effective Medium Treatment. *Proc. Natl. Acad. Sci. USA 87*, 15 (1990), 5856–5857.

Danksagung

Viele Menschen haben zum Gelingen dieser Arbeit beigetragen.
Mein Dank gilt zu allererst meinem Betreuer Ulrich Schwarz für die Anregung eines für mich hochinteressanten Themas und die hervorragende Betreuung. Neben der steten Unterstützung meiner Arbeit gab er mir die Möglichkeit, einen umfangreichen Überblick über die aktuellen Forschungsschwerpunkte in der Biophysik zu erhalten. Insbesondere der regelmäßige Kontakt mit aktiven Wissenschaftlern und die Ermöglichung der Teilnahme an interessanten Konferenzen waren dafür entscheidend.
Volkhard Helms war nicht nur an der Entwicklung der grundlegenden Idee des Projektes beteiligt, sondern half auch später mit hilfreichen Anregungen bei der Verwirklichung. Für die offenen und aufschlussreichen Diskussionen mit ihm und seinen Mitarbeitern Tihamér Geyer, Michael Hutter, Denitsa Alamanova und Ines Hörger bei den diversen Zusammenkünften in Saarbrücken und Heidelberg möchte ich mich bedanken.
Der gesamten Arbeitsgruppe “Schwarz” danke ich für die angenehme Atmosphäre. Durch die Vielfalt der bearbeiteten Themen konnte ich viel lernen und die angeregten Diskussionen während der Gruppentreffen haben mir immer viel Spaß gemacht. So trug der intensive Austausch mit Achim Besser, Benedikt Sabass, Carina Edwards, Christian Korn und Julian Weichsel sehr zu meiner Entwicklung bei. Gern erinnere ich mich zudem an die regelmäßigen gemeinsamen Unternehmungen wie die Grillabende und das Wichteln im Hause Schwarz, diverse Kinobesuche und sportliche Aktivitäten.
Weiterhin gilt mein Dank Achim Besser, Christian Korn und Julian Weichsel, die durch ihr sorgfältiges Korrekturlesen und viele durchdachte Vorschläge sehr zur Verbesserung des Manuskripts beigetragen haben.
Achim Besser chauffierte mich immer äußerst zuverlässig nach Karlsruhe. Durch die kurzweiligen Unterhaltungen gewann das Pendeln etwas Freizeitcharakter. Dafür vielen Dank.
Meiner Frau Isabel möchte ich für die schöne gemeinsame Zeit und ihre liebevolle Unterstützung besonders in den schwierigen Phasen der Promotion danken. Auch meinen Eltern, die mir stets warmherzigen Beistand geleistet haben, gilt mein Dank.
Nicht zuletzt danke ich der VolkswagenStiftung, die meine Arbeit im Rahmen der Förderinitiative “Neue konzeptionelle Ansätze zur Modellierung und Simulation komplexer Systeme” maßgeblich finanziert hat.