Fakultät für Physik
Technische Universität München

Physical Analysis of the Target Search of Surface Proteins for a Binding Site on a Confined Polymer

Linda Johanna Martini

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Friedrich C. Simmel

Prüfende der Dissertation:
1. Prof. Dr. Ulrich Gerland
2. Prof. Dr. Martin Zacharias

Die Dissertation wurde am 03.03.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 21.07.2022 angenommen.

Abstract

Biopolymers such as DNA or RNA are typically confined to a certain volume, like that of a cell or a nucleus, while a multitude of proteins interacting with DNA and RNA reside in the surrounding membrane. For several cellular processes it is hence of vital importance that a binding site on a biopolymer coincides with a particular molecule at the surface of the confining domain. Prime examples are mRNA molecules or DNA double-strand breaks relocating to nuclear pore complexes in eukaryotic cells and one-component signalling systems binding to a specific site on the DNA from the bacterial membrane. While the biological examples demonstrate that locating a specific binding site on a long polymer from the surface of the confining volume is feasible, the dynamics of these search processes are still highly uncharacterised. The striking question is how the specific polymer-protein contact is kinetically established. In contrast, the target search dynamics of two-component signalling systems featuring a sensor kinase and a separate cytoplasmic transcription factor are a well studied problem. The so-called facilitated diffusion model identified a combination of 3d diffusion in the cytoplasm and 1d sliding along the polymer as the underlying mechanism. To which extent is such an elaborate strategy required to locate a binding site on a polymer in confinement from the surface of the confining volume?

In this thesis we employ kinetic Monte Carlo simulations and mathematical modelling to analyse the search process between a confined polymer and a protein moving on the surface of the confining volume. After establishing a coarse-grained cubic lattice model that allows us to simulate the substantial time and length scales of the process, the kinetics are extensively analysed. For high densities of ideal polymers the numerical simulations reveal that due to the confinement the search process becomes independent of the polymer length, making the search feasible even for very long chains. In the opposite domain of low polymer densities transient tethering due to non-specific binding of the protein to the polymer emerges as a way to speed up the process. In contrast to the facilitated diffusion model for cytoplasmic transcription factors, sliding is less effective for proteins at the surface since the sliding range is largely limited to polymer segments close to the surface. The process is therefore mainly limited by diffusion of the protein and the polymer, respectively. When both diffusion rates are of similar order of magnitude, the underlying stochastic process can be described by a sequential two-step process. It reduces to a Poisson process with a single exponential distribution when either the polymer or the protein is approximately immobile. Overall, our findings reveal that elaborate strategies to speed up the procedure are mainly relevant at low polymer densities, while at high densities the decorrelation of polymer subchains due to the confinement facilitates the search process even for very long polymers.

In the last part of this dissertation we use a simplified model to simulate the target search of a membrane-integrated transcription factor in *Escherichia coli*. We focus on the case of CadC, the pH receptor of the acid stress response Cad system in *E. coli*. CadC is a prime example

of a one-component signalling protein that directly binds to its cognate target site on the chromosome to regulate transcription. Fluorescence microscopy experiments were combined with mathematical analysis and kinetic Monte Carlo simulations to probe this target search process. The time from activation of CadC until successful binding to the DNA in single cells was measured by labelling CadC with a fluorescent marker, which forms visible fluorescent spots as stable receptor-DNA complexes are formed. The experimental data suggest CadC to be highly mobile in the membrane, finding its target by a 2d diffusion and capture mechanism. The mean search time of CadC for its cognate binding site was measured between four and five minutes, unaffected by relocating the DNA binding site to a distant position on the chromosome. Using a mutant strain with two binding sites reduced the search time by about a factor of two. This behaviour is consistent with our numerical simulations of a coarse-grained lattice model for the coupled dynamics of DNA within a cell volume and proteins on its surface. Moreover, quantitative agreement between the numerically simulated and experimentally established distribution of search times was found. Overall, our findings reveal that the DNA target search does not present a much bigger kinetic challenge for membrane-integrated proteins than for cytoplasmic proteins. CadC diffusion in the membrane is pivotal for this search process, while the DNA binding site is just mobile enough to randomly reach the membrane.

Zusammenfassung

Biopolymere wie DNA und RNA sind typischerweise räumlich begrenzt auf ein bestimmtes Volumen, wie das einer Zelle oder eines Nukleus, während mit DNA und RNA interagierende Proteine häufig in der umschließenden Membran angesiedelt sind. Für viele zelluläre Prozesse ist es demnach unerlässlich, dass eine Bindestelle auf einem Polymer auf ein bestimmtes Molekül auf der Oberfläche des begrenzenden Volumens trifft. Zentrale Beispiele sind mRNA-Moleküle oder DNA-Doppelstrangbrüche, die sich in Eukaryoten zu Kernporen hin verlagern, sowie Einkomponenten-Signalmoleküle, die von der bakteriellen Membran aus an eine spezifische Bindestelle auf der DNA binden. Während die biologischen Beispiele demonstrieren, dass die Lokalisierung einer spezifischen Bindstelle auf einem langen Polymer von der Oberfläche des begrenzenden Volumens aus möglich ist, ist die Dynamik eines solchen Prozesses weitgehend unbekannt. Die zentrale Frage ist, wie der spezifische Kontakt zwischen Polymer und Protein kinetisch abläuft. Der Suchprozess von Zweikomponenten-Signalsystemen hingegen, die aus einer Sensorkinase und einem separaten zytoplasmischen Transkriptionsfaktor bestehen, wurde in der Vergangenheit eingehend untersucht. Das Modell der sogenannten erleichterten Diffusion (facilitated diffusion) konnte zeigen, dass der zugrundeliegende Mechanismus aus einer Kombination von dreidimensionaler Diffusion im Zytoplasma und eindimensionalem Gleiten entlang des Polymers besteht. Inwiefern bedarf es im Falle eines räumlich begrenzten Polymers einer solchen besonderen Strategie, um eine Bindestelle auf dem Polymer von der Oberfläche des begrenzenden Volumens aus zu lokalisieren?

Wir verwenden kinetische Monte Carlo Simulationen und mathematische Modellierung, um den Suchprozess zwischen räumlich begrenzten Polymeren und beweglichen Proteinen auf der Oberfläche des begrenzenden Volumens zu analysieren. Nach der Einführung eines grobkörnigen kubischen Gittermodells, das es uns erlaubt, die enormen Längen- und Zeitskalen des Prozesses zu simulieren, wird die Kinetik des Suchprozesses eingehend analysiert. Die Simulationen zeigen, dass für hohe Polymerdichten eines idealen Polymers die räumliche Begrenzung zu einer Unabhängigkeit des Suchprozesses von der Polymerlänge führt, sodass er auch für sehr lange Ketten realisierbar ist. Bei geringen Polymerdichten hingegen kann ein durch unspezifisches Binden des Proteins erzeugtes, vorübergehendes Anhaften des Polymers an der Oberfläche zu einer Beschleunigung des Verfahrens führen. Im Gegensatz zum Modell der erleichterten Diffusion für zytoplasmische Transkriptionsfaktoren spielt Gleiten entlang des Polymers für Proteine auf der Oberfläche eine untergeordnete Rolle, da die Distanz, die durch Gleiten zurückgelegt werden kann, durch die Polymersegmente nahe der Oberfläche stark begrenzt ist. Der Prozess wird daher hauptsächlich durch Diffusion des Proteins und des Polymers bestimmt. Wenn beide Diffusionsraten von gleicher Größenordnung sind, kann der zugrunde liegende stochastische Prozess durch einen sequentiellen Zweistufenprozess beschrieben werden. Dieser vereinfacht sich zu einem Poisson-Prozess mit exponentieller

Verteilung, wenn das Polymer oder das Protein statisch ist. Zusammenfassend zeigt unsere Analyse, dass besondere Strategien zur Beschleunigung des Suchprozesses hauptsächlich für geringe Polymerdichten relevant sind. Bei hohen Polymerdichten führt die Dekorrelation verschiedener Abschnitte des Polymers aufgrund der räumlichen Begrenzung dazu, dass der Prozess unabhängig von der Polymerlänge abläuft und damit auch für sehr lange Polymere durchführbar ist.

Im letzten Teil der Dissertation nutzen wir eine vereinfachte Version des Gittermodells, um den Suchprozess eines membranintegralen Transkriptionsfaktors in *Escherichia coli* zu simulieren. Das betreffende Protein ist CadC, der pH-Rezeptor des säureinduzierten Cad Systems in *E. coli*. CadC ist ein Einkomponenten-Signalprotein, das direkt an seine spezifische Bindestelle auf der DNA bindet, um Transkription zu regulieren. Wir nutzen fluoreszenzmikroskopische Experimente, mathematische Analyse und kinetische Monte Carlo Simulationen, um diesen Suchprozess zu erforschen. Durch Markieren von CadC mit einem fluoreszierenden Protein konnte die Zeit von seiner Aktivierung bis zum erfolgreichen Binden an die DNA in einzelnen Zellen gemessen werden, da das Bilden stabiler Rezeptor-DNA-Komplexe zu sichtbaren Fluoreszenzpunkten führt. Die experimentellen Daten demonstrieren die Beweglichkeit von CadC in der Membran, welches seine Bindestelle auf der DNA durch einen zweidimensionalen Suchmechanismus lokalisiert. Eine mittlere Suchzeit zwischen vier und fünf Minuten wurde gemessen, unabhängig von der Verschiebung der Bindestelle auf der DNA auf eine entfernte Position entlang des Chromosoms. Verwendung einer Mutante mit zwei Bindestellen hingegen führt zu einer Halbierung der mittleren Suchzeit. Dieses Verhalten ist konsistent mit unseren numerischen Simulationen zur gekoppelten Dynamik der DNA im Zellinneren und CadC auf der Oberfläche. Darüber hinaus konnte eine quantitative Übereinstimmung zwischen numerisch simulierten und experimentell bestimmten Verteilungen der Suchzeit gefunden werden. Zusammenfassend zeigen unsere Ergebnisse, dass der Suchprozess nach einer Bindestelle auf der DNA keine signifikant größere Herausforderung für membranintegrale Proteine darstellt als für zytoplasmische. Die Diffusion von CadC in der Membran ist ausschlaggebend für den Suchprozess, wohingegen die Bindestelle auf der DNA gerade mobil genug ist, um der Membran zufällig nahezukommen.

Contents

List of Figures

List of Tables

I Introduction

Life on earth, ranging from micron-sized bacteria and viruses over plants and fungi to highly complex organisms like animals and humans, is of overwhelming diversity and complexity. To this day, scientists have not even agreed on a coherent definition of life [1]. Despite the diversity of life on earth, there are some common principles and components that are highly conserved among all living systems. The basic unit of all known organisms is the cell, either constituting a unicellular organism of its own or forming a multicellular organism with up to 4×10^{13} highly specialised cells [2]. While different cells vary greatly in their shape, function and size, all cellular processes depend on the interplay of two basic families of biopolymers: polynucleotides, such as DNA and RNA, and proteins.

Proteins are involved in almost all tasks that ensure the proper functioning of living organisms: RNA polymerases transcribe DNA into RNA, DNA polymerases duplicate DNA, ribosomes function to synthesise proteins, enzymes act as catalysts and alter the rates of biochemical reactions. Moreover, proteins participate in the perception and transduction of extracellular signals, in light and sound detection, in the exertion of mechanical forces in muscles and much more [3]. While proteins are the workers of the cell, DNA is a macromolecule that stores all information relevant for the cell. Somewhere in between resides RNA, which can store information but also acts as an enzyme [4].

One of the most central forms of collaboration between polynucleotides and proteins emerges in transcriptional regulation and gene expression. Often triggered by proteins sensing a change in the environment, like nutrient concentrations or the acidity of the surrounding medium, proteins called transcription factors (TFs) can bind to specific sites on the DNA and either prevent or provoke transcription of a certain gene [4]. When transcription is induced, the concerning gene is copied into a strand of RNA (mRNA). In the nuclei of eukaryotes, the cells of plants, fungi and animals, the mRNA molecule again needs to interact with a protein, as it has to locate a so-called nuclear pore in order to exit the nucleus. In the cytoplasm of both eukaryotes and prokaryotes, the latter being the cells of bacteria and archaea, a ribosome, which in itself is a construct of RNA (rRNA) and proteins, translates mRNA into new proteins.

Besides the action of TFs, there exist many processes involving the interaction between polynucleotides and proteins that require the colocalisation of specific binding sites on the respective molecules with one another. In particular for proteins that are membrane-integrated and therefore highly constrained in their motion it seems like a challenging task to locate such a small strategic target on a polymer. Nevertheless, as there are many examples where cellular functions rely on these search mechanisms to succeed within reasonably short time,

it seems to be possible. An open question therefore is to which extent elaborate strategies are required in order to speed up the dynamics of these search processes. As for cytoplasmic proteins a mechanism called "facilitated diffusion" has been shown to be essential to achieve sufficiently short search times.

The objective of this thesis is to study the kinetics of target search processes in the cell. In particular, we are concerned with the question how a protein moving on the surface of a volume finds a binding site on a polymer inside the volume. We start in chapter II by developing an appropriate model that allows us to numerically simulate the search procedure. Chapter III covers an extensive analysis of the simulations, identifying various strategies to speed up the search process. In chapter IV the simulations are applied to compare to experiments on the target search of a membrane-integrated transcription factor in *Escherichia coli* to analyse the kinetics of this particular search mechanism.

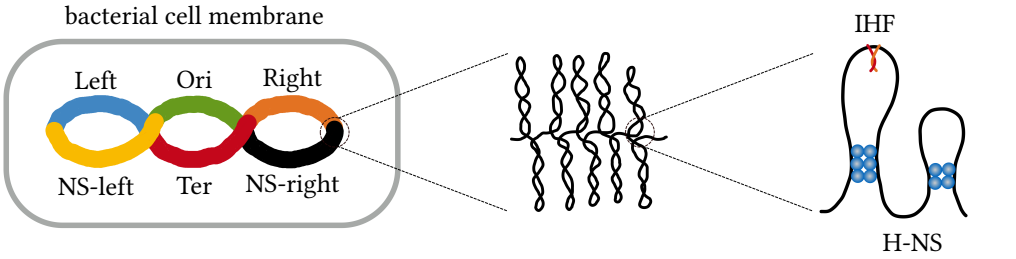
The following sections of this introductory chapter cover some aspects that are related or prerequisite to our work. First, we include a short discussion of the basics of transcriptional regulation, which is relevant to the main application of our simulations. Moreover, we give an overview of relevant search processes in the cell and a summary of previous studies investigating their kinetics. We conclude with a short introduction into polymer physics, covering some of the most relevant polymer models.

1 Gene Expression and Transcriptional Regulation

Most search processes considered in this thesis are connected to the mechanisms of gene expression and transcriptional regulation. We therefore start with a short introduction to how information is stored and processed in the cell. A more detailed introduction on this can be found in standard textbooks [4].

1.1 DNA and RNA

Deoxyribonucleic acid (DNA) is a linear polymer composed of four different monomeric groups, called nucleotides. Each nucleotide consists of a sugar called deoxyribose, a phosphate group and a base. There are two bases with a single aromatic ring (pyrimidines) called thymine (T) and cytosine (C) and two two-ring bases (purines) called adenine (A) and guanine (G). The nucleotides are connected by covalent bonds between the sugar of one nucleotide and the phosphate group of the next nucleotide, forming a sugar-phosphate backbone. The phosphate groups form phosphodiester bonds between the third (3′ end) and the fifth (5′ end) carbon atoms of two subsequent sugar groups. The backbone therefore has a directionality, starting with a deoxyribose at the 3′ end and ending with a phosphate group at the 5′ end. The bases of two such polynucleotide strands can bind via Watson-Crick base-pairing: adenine and thymine can bind via hydrogen bonds as well as guanine and cytosine. This means that DNA is constructed of two complementary strands, connected in an antiparallel fashion via hydrogen bonds. Since the bases are hydrophobic and the backbone is hydrophilic, the bases attract each other to exclude water molecules, while the distance between the sugar groups is fixed. The most efficient packing is therefore achieved by twisting the double-strand, leading to a double-helical structure of the DNA. Information is stored in the DNA in terms of a



(a) Macrodomains **(b)** Topological domains **(c)** nm-scale structure

Figure I.1: Hierarchical DNA organisation in *E. coli*. The bacterial chromosome is organised in a hierarchical way. (**a**) On the largest scale the chromosome is organised into macrodomains with a defined localisation within the cell. For example in *E. coli*, the four macrodomains are Ori, surrounding the origin of replication, Ter, surrounding the terminus-binding site, and Left and Right, which are on either side of the Ter macrodomain. On either side of the Ori macrodomain are the two unstructured domains NS-left and NS-right. (**b**) Negative supercoiling of the DNA leads to the formation of plectonemic loops, each of them forming a topological domain. (**c**) At the smallest scale, nucleoid associated proteins (NAPs) organise the chromosome by inducing bends, bridging of the DNA or wrapping the DNA similar to histones in eukaryotes. Examples in *E. coli* are the histone-like proteins IHF and H-NS that lead to DNA bending and bridging respectively [5].

four letter alphabet given by the four bases A, T, G and C. To access this information, the two strands have to open up, which is possible because the hydrogen bonds are less stable than the covalent bonds connecting the backbone. Once the double-strand is opened up, a protein called DNA polymerase can recreate the complementary sequence of each strand by Watson-Crick base-pairing. This process is called replication and yields two identical copies of double-stranded DNA (dsDNA) which are distributed among the two daughter cells during cell division.

Ribonucleic acid (RNA) has a similar structure as single-stranded DNA (ssDNA). The only differences are the sugar group, which consists of ribose instead of deoxyribose, and the base thymine, which is replaced by uracil (U) in the case of RNA.

1.2 Genetic Code

The sequence of bases in the DNA stores all information relevant for the cell. It contains sections (genes) that code for certain proteins or regulatory RNA molecules, as well as regulatory sequences that control where, when and how many proteins or RNA molecules are produced. Especially in prokaryotes genes are often organised in so-called operons, a cluster of genes under the control of a single promoter. How the four letter alphabet of the DNA (or RNA) is translated into a sequence of 20 (22) different amino acids is determined by the genetic code. Three bases form a so-called codon which codes for a specific amino acid.

The process of producing proteins from a certain DNA sequence starts with transcription. The DNA double-strand is opened up, such that a protein called RNA polymerase can bind to the promoter, directly upstream of the gene. It moves towards the 3′ end and copies

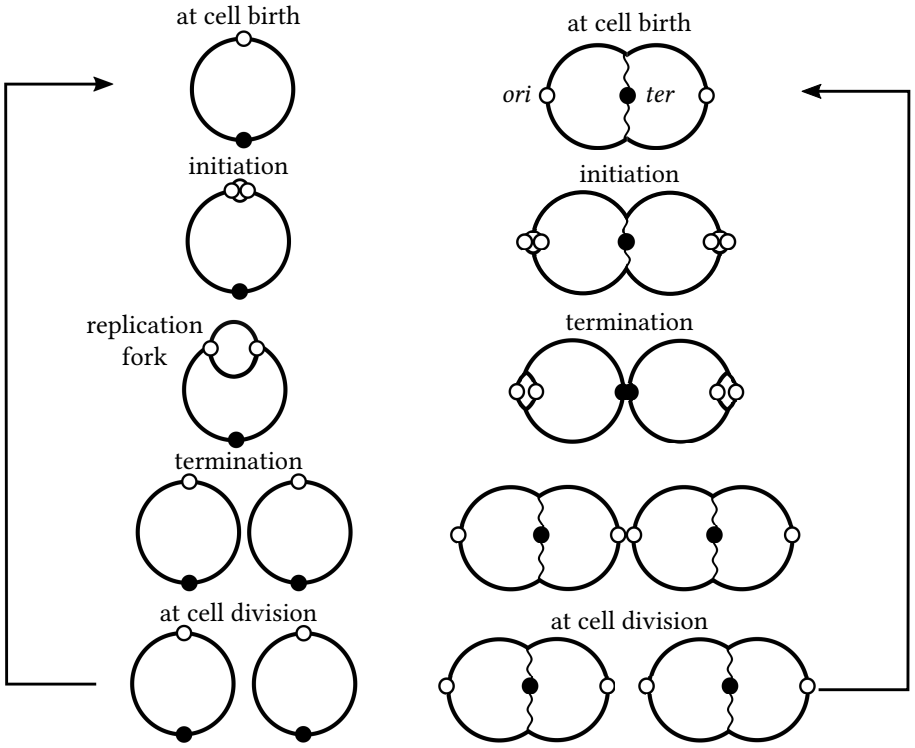
one DNA strand into a messenger RNA (mRNA) molecule. In a process called translation the mRNA sequence is used by a ribosome, a large complex of proteins and ribosomal RNA (rRNA), to create a chain of amino acids. The mRNA molecule is pulled through the ribosome, three nucleotides at a time, and fitted to an anticodon (triplet of complementary nucleotides), to which the respective amino acid is bound. When the "stop" codon is reached, the newly synthesised polypeptide, i.e. a chain of amino acids is released. One or multiple polypeptides fold into a specific 3d structure, forming a protein, the structure being an important determinant for its function.

1.3 DNA Organisation and Replication

Both in eukaryotes and in prokaryotes the spatial organisation of the genome is a highly complex subject. DNA has to be compacted by a factor of 1000–7000, while still ensuring its accessibility for processes like replication, transcription and DNA repair.

In eukaryotes the DNA is contained in a compartment called the nucleus, surrounded by a membrane. The eukaryotic genome is divided into multiple chromosomes, each of them containing a very long DNA molecule. When stretched out, for example the 46 chromosomes of the human genome have a total length of about 2 m, which need to be compacted heavily to fit into the nucleus with a diameter of 6 µm. On the lowest level the DNA composing each chromosome is wrapped 1.7 times around octamers of proteins called histones, forming a nucleosome. The complex of DNA and histones is called chromatin, which forms a "bead on a string" structure with linker DNA in between the nucleosomes. The nucleosomes are then tightly packed, forming a fibre of about 30 nm in diameter. Supercoiling of the 30 nm fibre leads to a bottlebrush structure, the most compact form of eukaryotic DNA. The packing of DNA is highly dynamic, which is important for the accessibility of the stored information and dependent on the stage in the cell cycle.

Prokaryotes have no specific internal structure like eukaryotes. The circular bacterial chromosome replicates bidirectionally starting from the origin of replication (*ori*) and ending at the terminus-binding site (*ter*). The position of a chromosomal locus along the bacterial DNA is commonly described in units of minutes, ranging from zero to 100 in case of *E. coli*. This refers to the time it takes until the concerning locus is transferred from a donor cell to a recipient cell in a conjugation process, which was used to obtain this mapping, such that in *E. coli ori* is located at 84.2′ and *ter* is located at 33.7′ [6]. *E. coli* is a rod shaped bacillus, that is about 1 µm long and 0.35 µm wide and is a constituent of the mammalian gut microbiome, where it produces vitamin K and vitamin B12 and excludes pathogens from its niche in the gut [7]. Since it can be grown and cultured easily *E. coli* has become the most commonly studied prokaryotic model organism. Pathogenic strains however are a major cause of diarrheal diseases, urinary tract infections, meningitis, Crohn's disease and many more [8]. The prokaryotic genome, consisting mostly of a single circular chromosome, was long thought to be completely unstructured. However, recent studies have shown that it is organised in a hierarchical fashion inside a nucleus-like body, called the nucleoid [9, 10]. As the nucleoid is not surrounded by a membrane, the bacterial chromosome is compacted solely due to interactions with organising proteins, as depicted in fig. I.1c. On the smallest scale, negative supercoiling of the DNA leads to the formation of plectonemic loops, each of



(a) Slowest growth **(b)** Faster growth

Figure I.2: Chromosome organisation in *E. coli*. The cell cycle dependent organisation of the chromosome in *E. coli* is shown. The origin of replication (*ori*) is depicted as a black circle, the terminus (*ter*) is represented by a filled black circle. The long axis of the cell stretches from left to right. (**a**) At slowest growth conditions the *E. coli* chromosome obtains a Left-*ori*-Right configuration. (**b**) At faster growth conditions the cell cycles overlap and the multifork chromosome in *E. coli* is in *ori-ter-ori* configuration.

them forming a topological domain [9, 10], shown in fig. I.1b. The loops are thought to be held together by small proteins, called domainins, that are concentrated at the nucleoid core, thus leading to a bottlebrush structure of the chromosome. A large variety of DNA-binding proteins have been identified to be involved in compacting the bacterial DNA by inducing bends, bridging of the DNA or wrapping the DNA similar to histones in eukaryotes [11, 10]. On a larger scale the bacterial chromosome is organised into macrodomains, characterised by chromosomal loci that frequently co-occupy the same cytoplasmic space, as depicted in fig. I.1a. For example in *E. coli*, two of the four macrodomains that have been identified are surrounding the origin of replication and the terminus-binding site, respectively, called the Ori and Ter macrodomains. Next to the Ter macrodomain are the Left and Right macrodomains, while the Ori macrodomain is surrounded by two unstructured domains NS-left and NS-right.

There are two general ways of spatial organisation of bacterial chromosomes [12]. The most common organisation is the longitudinal *ori-ter* configuration, with the origin at the old

pole and the terminus at the new pole of the cell. However, some bacteria, like *E. coli* at slowest growth conditions, are organised in a transverse fashion, in the Left-*ori*-Right configuration with the origin and the terminus at the centre along the long axis of the cell, as shown in fig. I.2a. After replication the origins are segregated to the cell quarter positions and the newly replicated Left and Right arms segregate to either side to recreate the transverse configuration in the next generation. In contrast to eukaryotes, where replication and segregation are clearly separated in time, it has been shown that in *E. coli* the cell cycles overlap for all but the slowest growth conditions [6]. At faster growth conditions the so-called multifork chromosome adopts a different organisation with the origin at the cell poles and the terminus at the centre, as demonstrated in fig. I.2b.

1.4 Proteins

Proteins are the molecules that are the most prevalent in all cells and are involved in all types of cellular functions. Examples include enzymes like DNA polymerase, transport proteins like haemoglobin, effector proteins like insulin, receptor proteins, transcription factors and many more. Proteins are large biomolecules, that are constructed from one or multiple polypeptide chains. Polypeptides are linear polymers composed of 20 different amino acids (22 including selenocysteine and pyrrolysine that are included in certain organisms), connected by peptide bonds between the amino group of one and the hydroxyl group of the next amino acid. A polypeptide therefore starts with an amino group at the N-terminal end and ends with a hydroxyl group at the C-terminal end. Polypeptide chains fold in a very specific way, determining the function and binding properties of the respective protein. Ligands binding to a protein can change its conformation reversibly, thereby influencing its binding properties, leading to activation or inactivation of the protein. The two main folding motifs are β-sheets and α-helices, the latter being a common structure of transmembrane proteins, stretching through the membrane in a helical shape.

1.5 Transcription Factors

There are many ways how a cell can control and regulate its proteome. Transcriptional regulation being the most important for most genes, control of protein lifetime and translation can play a role, as well as the regulation of splicing and mRNA localisation in eukaryotes.

Transcription factors play a central part in gene regulation and are among the largest protein classes in cells. They are proteins that directly bind to a specific site on the DNA to either activate or repress transcription of a certain gene. In bacteria, most transcription factors bind close to the promoter and either interact with RNA polymerase to activate transcription or occupy the promoter to repress transcription. Some bind further away and influence transcription by inducing loops in the DNA that either bring the RNA polymerase closer to the promoter or hide the promoter and therefore suppress transcription.

Many transcription factors are controlled by other proteins. A common mechanism in bacteria are signalling systems, where a membrane-integrated sensor kinase registers a change in the environment and either activates or deactivates a response regulator. A well-known example concerns the control of the *lac* gene. When the surrounding medium provides a

sufficient amount of glucose, it is the only food source of *E. coli* [13]. A lack of glucose in the environment however leads to the metabolism of other sugars like lactose. The switch in the metabolic pathway is controlled by the catabolite activator protein (CAP) and the Lac repressor. The CAP activator is a universal transcriptional activator, that binds to the DNA and facilitates binding of RNA polymerase to increase the level of transcription of the lactose metabolising proteins when the glucose supplies become exhausted [14]. The Lac repressor however binds two distal sites on the chromosome and induces bending of the DNA in between to prevent transcription [13]. It is only deactivated when lactose is available, leading to the desired result that lactose metabolising proteins are only expressed when they are required to keep the cell alive, i.e. when there is no glucose but enough lactose to be taken up in the environment.

1.6 Binding Kinetics of TFs

As stated above, transcription factors function in gene regulation by directly binding to the DNA. Fast and reliable regulation of gene expression requires fast search and recognition of the specific binding site by the transcription factor, as well as stable binding of the protein-DNA complex. In *E. coli* the genome contains about 3.60×10^6 bp (basepairs), whereas a specific binding site is about 5–10 nucleotides long. For a protein P to bind reversibly to a specific site on the DNA therefore requires that the protein first locates this binding site by diffusion. Especially for low copy number transcription factors like the Lac repressor, with a very limited number of molecules per cell, this can constitute the main limiting step in the binding process. The second step is to bind to the cognate binding site via chemical reaction.

The chemical reaction scheme is given by

$$P + \mathrm{DNA} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} P|\mathrm{DNA}\,,$$

where $P|\mathrm{DNA}$ is the protein-DNA complex, k_{on} is the effective binding rate and k_{off} is the effective unbinding rate. In steady-state, the concentration $[\cdot]$ of the protein-DNA complex changes with time according to the steady-state rate equation

$$\frac{\mathrm{d}}{\mathrm{d}t}[P|\mathrm{DNA}] = k_{\mathrm{on}}[P][\mathrm{DNA}] - k_{\mathrm{off}}[P|\mathrm{DNA}]\,.$$

At equilibrium the concentrations are constant and it follows

$$\frac{[P][\mathrm{DNA}]}{[P|\mathrm{DNA}]} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}}\,,$$

which is known as the law of mass action [15] and defines the equilibrium constant $K_{\mathrm{eq}} = k_{\mathrm{off}}/k_{\mathrm{on}}$. The law of mass action is widely used to treat chemical reactions as it provides a way to compute the equilibrium constant by measuring particle concentrations in a test tube. From such experiments it can be directly shown how binding of a protein to DNA depends on whether the DNA strand contains a specific binding site or not. The standard free energy change for specific binding is of the order of $20–25 k_{\mathrm{B}}T$ and for non-specific binding $5–10 k_{\mathrm{B}}T$

[3]. Note that $10k_BT$ is the energy delivered by hydrolysis of an ATP molecule, which is the energy currency of the cell.

The steady-state reaction rates depend both on diffusion and on the intrinsic reaction rates $\tilde{k}_{on}$ and $\tilde{k}_{off}$ and are given by

$$k_{on} = \frac{k_D \tilde{k}_{on}}{k_D + \tilde{k}_{on}},$$
$$k_{off} = \frac{k_D \tilde{k}_{off}}{k_D + \tilde{k}_{on}},$$

with the Smoluchowski rate $k_D = 4\pi(D_{DNA} + D_P)R$, the diffusion constants of the two molecules D_{DNA} and D_P and the contact radius R, i.e. the sum of the radii of the two reactants. The Smoluchowski rate is obtained by solving the steady-state diffusion equation for the concentration of P with the absorbing boundary condition $[P](R) = 0$ and assuming that the protein concentration reaches its bulk value for $R \to \infty$ [16]. This means that the two species are treated as spherical Brownian particles that react as soon as they come into close contact.

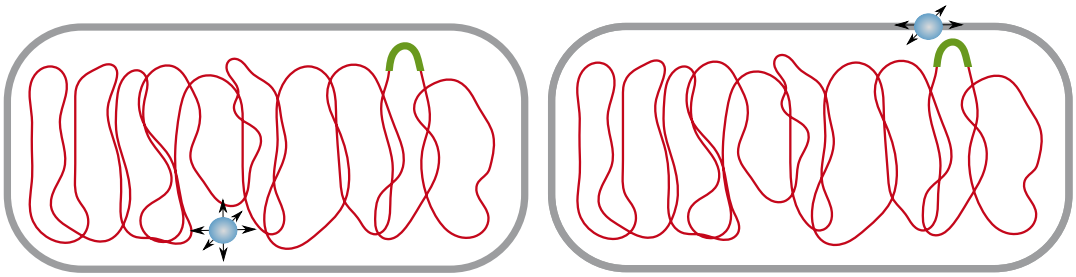
The effective steady-state reaction rates depend on the diffusion constants of the two species via the Smoluchowski rate and on the intrinsic reaction rates. For large intrinsic association rates $\tilde{k}_{on} \gg k_D$ the reaction is said to be "diffusion limited". In this regime perfect reaction is assumed and the effective association rate reduces to the Smoluchowski rate $k_{on} = k_D$. No bimolecular rate constant in 3d involving a diffusive encounter without attractive forces can exceed this diffusion-controlled reaction rate. The opposite regime $k_{on} \ll k_D$ is called "reaction limited". In this case, diffusion is so fast that it perfectly mixes the two species and the effective association rate becomes $k_{on} = \tilde{k}_{on}$ [17].

In order to bind to the DNA and function in gene regulation a transcription factor therefore faces a search process to locate the cognate binding site by diffusion. There are many other examples of similar search processes in eukaryotic and prokaryotic cells, some of which will be introduced in the next section.

2 Search Processes in the Cell

Many processes in the cell are based on the encounter of a protein with a specific binding site on a polymer. These search processes include the initiation of DNA replication, RNA polymerase binding to the promoter during gene expression, ribosomes searching for mRNA and much more. In the context of gene delivery, DNA fragments have to locate nuclear pores to enter the nucleus and the repair process of DNA double-strand breaks requires the colocalisation of the two free ends.

It is of vital importance to bacteria to be able to rapidly adapt to changing environmental conditions like nutrient composition, environmental stresses and antibiotics. This is enabled by sophisticated signalling schemes, primarily based on one- and two-component systems [18, 19]. Two-component signalling systems feature a sensor kinase and a separate response regulator, where the former is typically membrane-integrated while the latter diffuses through the cytoplasm to reach its regulatory target [19]. The majority of response regulators are transcription factors that bind to specific target sites on the genomic DNA to activate or repress



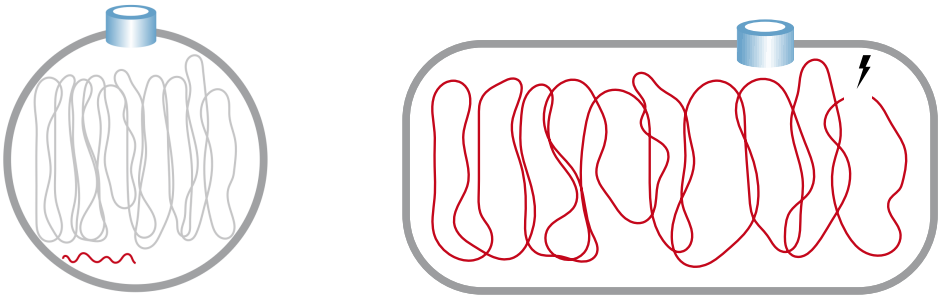
(a) Cytoplasmic transcription factor **(b)** Membrane-integrated transcription factor

Figure I.3: DNA target search of cytoplasmic vs. membrane proteins. (**a**) A cytoplasmic transcription factor (blue) locates its specific binding site (green) on the DNA (red) through a combination of 3d diffusion in the cytoplasm and 1d sliding along the polymer. While the DNA itself is not static during this target search, DNA motion does not significantly contribute to its completion, since the search by the diffusing protein is faster and more efficient. (**b**) In contrast, a membrane-integrated transcription factor (blue) can only perform 2d diffusion in the membrane (grey), so that DNA motion (red) becomes essential. At a minimum, the specific DNA binding site (green) has to move close to the membrane to enable target recognition.

transcription. Hence, a key step in the signal transduction pathway of these two-component systems is a DNA target search by a cytoplasmic protein, shown in fig. I.3a.

In one-component signalling systems however, a single protein combines both sensory function and response regulation [18]. These one-component systems that are both membrane-integrated sensors and DNA-binding response regulators therefore face an extraordinary DNA target search problem: They must locate and bind to a specific site on the bacterial chromosome from the membrane, depicted in Figure I.3b. The ToxR receptor family is an example of such one-component systems. They comprise a periplasmic sensory domain and a single transmembrane helix that is connected via a linker to a cytoplasmic DNA-binding domain [20]. ToxR in *Vibrio cholerae* is a transcriptional activator that controls transcription of cholera toxin, a major virulence factor for this pathogen [20]. Besides ToxR in *V. cholerae*, members of this receptor family include TcpP and TfoS in *V. cholerae* [21], PsaE in *Yersinia tuberculosis* [22], WmpR in *Pseudoalteromonas tunicata* [23] and the pH-stress-sensing receptor CadC in *E. coli* [24] and *V. cholerae* [25]. Also in eukaryotes there are membrane-located proteins that interact with the DNA. An example is the membrane protein Yhm2p in *Saccharomyces cerevisiae*, which is integrated in the mitochondrial membrane and directly binds to the mitochondrial DNA [26]. Yhm2p is a mitochondrial carrier that exports citrate from and imports oxoglutarate into the mitochondrion. Oxoglutarate and citrate are two key intermediates in the citric acid cycle [27] (tricarboxylic acid cycle, Krebs cycle) and the major source of cellular ATP production [28].

In eukaryotes, a prime example of a membrane protein interacting with DNA and RNA is the nuclear pore complex (NPC), a large assembly constructed from about 30 different proteins (nucleoporins) [29]. In order to be translated, mRNA molecules have to locate nuclear pores and exit from the nucleus [30], as shown in Figure I.4a. Besides their role in macromolecular transport in and out of the nucleus, recent experiments have shown nuclear pores to be



(a) mRNA escape **(b)** Double-strand break

Figure I.4: Polynucleotides migrating towards NPCs. A prime example of a membrane protein interacting with polynucleotides such as DNA and RNA are nuclear pore complexes (NPCs), shown in blue. (**a**) In order to be translated in the cytoplasm, mRNA molecules (red) first have to exit the nucleus through a nuclear pore. (**b**) Persistent double-strand breaks in the DNA (red) are recruited towards NPCs, which has been shown to facilitate the repair process.

targeted by sites on the chromosomal DNA. While many forms of spontaneous DNA damage can be repaired within minutes, others appear to be particularly difficult and persist for hours. For some of those persistent lesions, relocation to NPCs, depicted in fig. I.4b, has been shown to facilitate the repair process [31]. Real-time imaging of live yeast cells has demonstrated recruitment of a fluorescently tagged double-strand break to nuclear pores [32]. Binding of genes at nuclear pores has also been shown to play a role in transcriptional regulation. According to the gene gating hypothesis [33], transcriptionally active genes are brought next to NPCs to direct export of transcripts through the pore and facilitate transcriptional regulation. Evidence for the gene gating hypothesis has been found for *S. cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and mammalian model systems [34]. Indeed, some active genes interact with components of NPCs and binding to the NPC can be induced by environmental stimuli such as nutrient shifts or heat shock [35]. In yeast, NPC interaction has been shown to be required for proper activation of certain genes, as tethering to NPCs promotes stronger expression and blocking NPC interaction reduces expression. NPC association appears to be guided by promoter DNA elements and requires specific DNA sequences, termed DNA zip codes [35].

Depending on spatial constraints and kinetic properties of searcher and target the cellular search processes can be described by different models, some of which will be introduced in the following paragraphs.

2.1 Cytoplasmic TFs

The target search dynamics of cytoplasmic transcription factors first started to raise attention in 1970, when Riggs et al. measured an *in vitro* association rate of the Lac repressor in *E. coli* of $k_a = 7 \times 10^9\,\mathrm{M^{-1}\,s^{-1}}$ [36]. Computation of the Smoluchowski rate for the Lac repressor locating its binding site on the DNA yields a value that is two orders of magnitude smaller

than the experimental measurement [37]. The Smoluchowski rate defining the upper limit for diffusion-controlled reactions, it could be concluded that the respective search process cannot proceed by pure 3d search of the protein. To resolve this discrepancy, a mechanism called "facilitated diffusion" was assumed to speed up the process. Inspired by the idea that a reduction of dimensionality can lead to enhanced reaction rates [38], Richter and Eigen suggested that the transcription factors combine 3d diffusion in the cytoplasm with 1d sliding along the DNA molecule, while being non-specifically bound. As non-specific binding was shown to be driven mainly by screened electrostatic interactions between charged DNA and protein molecules [39], it is highly dependent on salt concentration. The combination of 3d search and sliding along the polymer could therefore successfully explain both the high association rates and the dependence on the ionic strength measured in experiments [40]. This idea was extended to a comprehensive kinetic theory for the search process of cytoplasmic proteins by Berg, Winter and von Hippel [41, 42, 43], setting the basis of ongoing theoretical studies on facilitated diffusion until today. The corresponding *in vitro* experiments identified an optimum in the association rate depending on the salt concentration, arising due to the redundancy of the 1d search [44]. When the protein spends too little time sliding, very few sites are scanned and a large number of such rounds is needed to locate the target. When it spends a long time sliding however it scans the same sites multiple times and the search becomes very inefficient.

An intuitive approach to facilitated diffusion is presented in [37]. We assume a single protein that locates a single target site on the polymer of length N by alternating rounds of 3d diffusion and 1d sliding along the polymer. With τ_{3d} and τ_{1d} the average time spent in the respective search modes and $\bar{k}$ the mean number of rounds needed until the target site is found, the mean search time is $\tau = \bar{k}(\tau_{3d} + \tau_{1d})$. When the protein scans on average $\bar{n} \ll N$ sites on the polymer, the probability to locate the target in a single round of sliding is $p = \bar{n}/N$. When the intervals of 3d diffusion lead the protein to reassociate at a random position along the polymer, the sliding events are independent. The probability that the first encounter with the target requires k trials is therefore given by the geometric distribution

$$\mathcal{P}(X = k) = (1 - p)^{k-1} p\,,$$

for $k \in \{1, 2, \dots\}$. The expectation value of the geometric distribution is $\bar{k} = 1/p = N/\bar{n}$. For normal diffusion of the protein along the polymer it follows that $\bar{n} \propto \sqrt{D_{1d}\tau_{1d}}$, which can be used to rewrite the mean search time and compute the derivative

$$\begin{aligned}\frac{\mathrm{d}\tau}{\mathrm{d}\tau_{1d}} &= \frac{\mathrm{d}}{\mathrm{d}\tau_{1d}} \frac{N}{\sqrt{D_{1d}\tau_{1d}}} (\tau_{3d} + \tau_{1d}) \\ &= \frac{N}{2\sqrt{D_{1d}\tau_{1d}}} \left(1 - \frac{\tau_{3d}}{\tau_{1d}}\right) .\end{aligned}$$

At the minimum search time $\mathrm{d}\tau/\mathrm{d}\tau_{1d} = 0$ we obtain $\tau_{3d} = \tau_{1d}$. The theoretical interpretation of the experimentally measured optimal association rate in dependence on the ion concentration is therefore, that the search is optimal when the protein spends the same time in both search modes. The speed-up by facilitated diffusion is hence achieved by effectively enlarging the target size to $\bar{n}$ sites on the polymer ("antenna" effect) [45, 37] as compared to pure 3d diffusion.

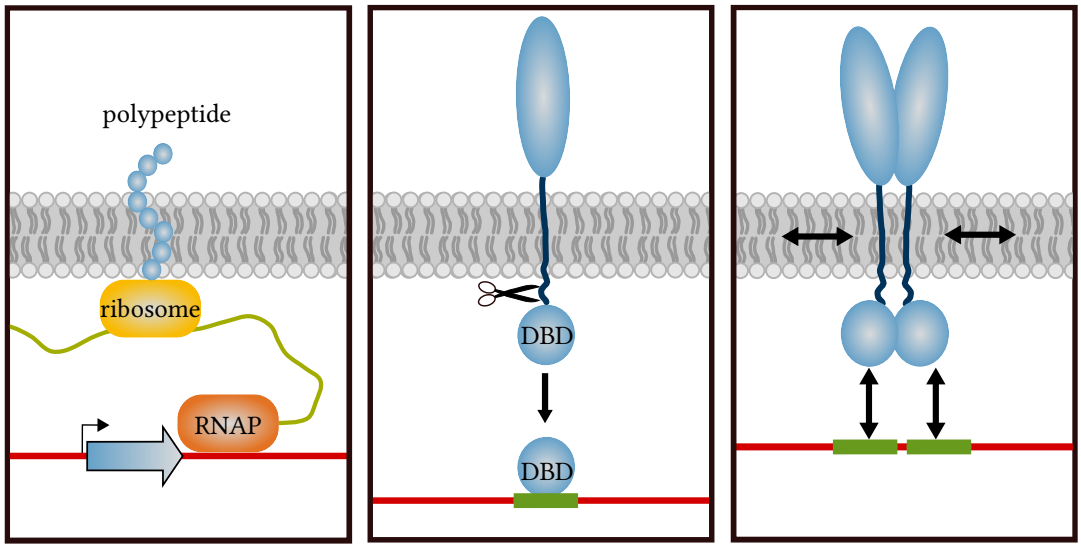
Besides one-dimensional sliding along the polymer, the theory of facilitated diffusion includes several other mechanisms of how a protein locates its binding site on the DNA [41, 42, 43]. One of them is called hopping or jumping along the polymer, where either microscopic dissociation leads to reassociation at a nearby DNA site or macroscopic dissociation leads to rebinding at a distal location. Another mechanism is intersegment transfer, where DNA looping enables the protein to transiently bind the two DNA strands at the intersection of the loop and move to the remote position. However, most fluorescence and microscopy based experiments could not distinguish between these processes and facilitated diffusion was therefore mainly ascribed to one-dimensional sliding along the polymer. Recent *in vitro* experiments on the restriction enzyme BbvCI showed that 1d diffusion is limited to short distances, while dissociation and reassociation was identified as the main mode of translocation [46]. A direct observation of sliding and jumping has been reported using fluorescence microscopy of the EcoRV restriction enzyme [47]. Furthermore, the impact of intersegment transfers on the association rate of EcoRV was shown by optical tweezers experiments, that show the effect the DNA conformation has on the search process [48]. In eukaryotes, single-molecule microscopy of the *S. cerevisiae* protein complex Mlh1-Pms1 could demonstrate how disruption of sliding along the DNA by hopping is used to bypass nucleosomes [49]. Moreover, the advancement of single-molecule techniques enabled the demonstration of sliding and 3d translocation also in living bacteria [50, 51].

The most controversial observation related to the facilitated diffusion model is called the "speed-selectivity paradox" [44]. It states that while the experimentally measured fast association rates suggest a smooth protein-DNA binding potential, the stability of the specific protein-DNA complex requires a large energy gap. However, recent analytical considerations together with Monte Carlo simulations resolve this paradox, which might arise due to the usage of continuum theoretical models beyond their range of validity [52]. An alternative solution is a two-state process, where the conformation of the protein switches between a search state and a recognition state [44, 53], supported by *in vitro* experiments of TALE protein dynamics [54].

The advancement of modern computers enabled the investigation of facilitated diffusion in molecular dynamics [55] and Brownian dynamics simulations [56], thereby bridging the gap between the mesoscopic picture and microscopic rates. Further studies continued to add to the detailed understanding of this target search process, e.g. with respect to effects of DNA conformation [57], DNA dynamics [58], and macromolecular crowding [59, 60].

2.2 Narrow Escape

Besides the search dynamics of proteins diffusing in the cytoplasm discussed above, a well studied problem is the mean first passage time (MFPT) of a Brownian particle to reach a small absorbing window in the otherwise reflecting boundary of a bounded domain. This so-called narrow escape problem has originated in the theory of sound [61, 62], but there has been renewed interest due to its relevance to many biological problems. Examples include ions diffusing inside a biological cell escaping through a protein channel in the membrane and mRNA molecules exiting the nucleus.



(a) Transertion **(b)** Proteolytic cleavage **(c)** Diffusion and capture

Figure I.5: Localisation models of membrane-integrated proteins. (**a**) The transertion mechanism describes simultaneous transcription, translation and membrane-insertion of the membrane protein. The DNA is shown in red, with the transcribed gene represented by the blue arrow. RNA polymerase (RNAP) transcribes the gene into mRNA (green), which is translated by a ribosome (yellow) into a polypeptide chain (blue). (**b**) The DNA-binding domain (DBD) of a membrane protein is proteolytically cleaved to search for its specific binding site (green) on the DNA as a regular cytoplasmic transcription factor. (**c**) In the diffusion and capture mechanism the whole protein stays membrane bound and locates its specific binding site (green) by two-dimensional diffusion in the membrane.

For a Brownian particle with diffusion constant D, the mean search time (narrow escape time, NET) to reach a circular target of radius a_t at the boundary of a spherical domain with volume V and radius R is given by the approximate expansion [63]

$$\tau = \frac{V}{4a_t D}\left(1 + \frac{a_t}{R}\log\left[\frac{R}{a_t}\right] + O\left(\frac{a_t}{R}\right)\right). \tag{I.1}$$

In first order, i.e. $R \gg a_t$ it reduces to Lord Rayleigh's formula $\tau \approx V/(4Da_t)$ [62, 64], which is valid for a circular target in a general 3d volume [63].

A similar estimate can be established for membrane proteins performing 2d diffusion towards a target, like the membrane-integrated proteins CadC in *E. coli* moving towards each other to form a homodimer [65]. For a Brownian particle with uniform initial distribution moving on the surface of a spherical domain of radius R, the mean first passage time to reach a circular target of radius a_t is approximately given by [66]

$$\tau = \frac{R^2}{D}\left(2\log\left[\frac{R}{a_t}\right] + 2\log 2 - 1 + O\left(\frac{R^2}{a_t^2}\log\left[\frac{R}{a_t}\right]\right)\right). \tag{I.2}$$

More asymptotic approximations of the MFPT are available in 2d [67, 68] and in 3d for elliptical and circular holes [64] and inside a sphere [69], which correspond to other biological scenarios of interest.

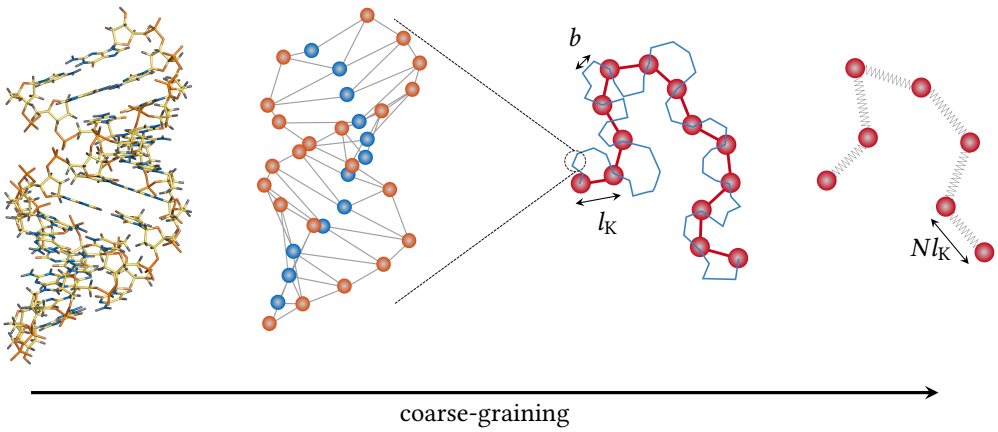
2.3 Membrane Proteins Locating a Binding Site on a Polymer

While the search dynamics of freely diffusing particles for a specific binding site on a polymer or a target at the surface of the confining volume have been extensively studied, little is known about how surface bound proteins locate a binding site on a polymer diffusing in the volume. Nevertheless, as stated above, many processes in bacterial and eukaryotic cells rely on such a search to proceed within reasonable time.

Three general pathways of membrane proteins localising a specific binding site in bacterial cells or eukaryotic nuclei are depicted in fig. I.5. Simultaneous transcription, translation and membrane-insertion of the protein ("transertion" [70, 71]), shown in fig. I.5a, lead to a tethering of the DNA locus of the one-component signalling system to the membrane. Since the specific binding site is typically close to the gene encoding the one-component system, this would bring the membrane protein close to its binding site on the DNA. Proteolytic cleavage of the DNA-binding domain of the membrane protein, demonstrated in fig. I.5b, leads to a separation from the rest of the molecule, such that it can search for its specific binding site as a regular cytoplasmic transcription factor. The third option is a diffusion and capture mechanism [72], demonstrated in fig. I.5c, where the membrane protein locates its cognate binding site by diffusion within the membrane and conformational fluctuations of the chromosome, bringing the binding site closer to the membrane.

There has been experimental evidence for all three mechanisms of localisation in bacterial cells. Transertion has been suggested to play a role in chromosome segregation in bacteria [73] and is regarded the main cause of large-scale heterogeneity in bacterial membranes [74]. Experimental evidence of the transertion mechanism is mainly based on the observation of genes of membrane proteins being drawn towards the membrane upon activation of expression [75] and that drugs inhibiting translation elongation lead to contraction of the nucleoid [76]. Proteolytic cleavage is a very common mechanism in both generating and receiving signals, that is found in eukaryotes and prokaryotes [77]. For example the acid-sensing membrane-integrated transcription factor CadC in *Salmonella enterica* was shown to be activated by proteolytic cleavage of the cytoplasmic domain under acid stress conditions [78]. Diffusion and capture is a widespread mechanism in all bacterial cells [79]. It was directly observed in *Bacillus subtilis* by fluorescence microscopy of a membrane protein (SpoIVFB) fused to a green fluorescent protein (GFP). SpoIVFB was found to be randomly positioned in the cytoplasmic membrane and to accumulate at the septum during initiation of sporulation, suggesting a mechanism of random membrane insertion followed by 2d diffusion and capture at the target location [80].

For the ToxR-like one-component receptor CadC in *E. coli*, experimental evidence has led to discard both proteolytic cleavage of the DNA-binding domain and a transertion mechanism in favour of the diffusion and capture model [81, 82, 83]. However, to the best of our knowledge there has not been any direct measurement of the dynamics of membrane-integrated proteins locating their binding site on the chromosome.



(a) Atomistic model | **(b)** United atom model | **(c)** Freely jointed chain | **(d)** Gaussian chain

Figure I.6: Polymer models. DNA can be modelled by a multitude of different polymer models, depending on the desired amount of detail. (**a**) At the highest resolution DNA is represented by an atomistic model. (**b**) Grouping multiple atoms together to form a monomer, DNA can be described on a coarser level by a united atom model. (**c**) At lower resolution the polymer is represented by a linear chain (blue) with arbitrary bond length b, which can be approximated by a freely jointed chain (red) with Kuhn segments of length l_K. (**d**) Many (N) Kuhn segments can be approximated by a Gaussian bead-spring chain with mean segment length Nl_K.

Also from the theoretical and computational side, previous work studying the encounter dynamics of a surface located particle with a target on the enclosed polymer is rare. Mathematical analysis and Langevin dynamics simulations of the end monomers of a polymer confined in a spherical domain have been used to compute the MFPT of reaching a nanopore at the boundary [84]. The main result is the scaling of the MFPT ("attempt time") τ with the confinement radius R of $\tau \propto R^{3.67}$ for flexible chains and $\tau \propto R^{2.67}$ for stiff chains. In a similar approach Brownian dynamics simulations have been used to simulate a polymer in a spherical volume where either one or all monomers can bind to a nuclear pore at the surface [85]. When only one monomer was allowed to bind to the target a strong dependence of the MFPT on its location along the polymer and on polymer length was found. In [86] pruned-enriched-Rosenbluth (PERM) simulations, i.e. a Monte Carlo method with importance sampling has been used to estimate the change in the internal conformational free energy upon localisation of an end monomer to either an infinite hard wall or the corner of an infinitely large cube.

While these previous studies focus on the encounter of a polymer site with a static molecule at the boundary, membrane proteins like channels, receptors and transporters are known to be highly mobile in the membrane [87]. In this thesis we aim at extending the discussion of the search dynamics to include proteins diffusing on the surface of the confinement. Moreover

we incorporate effects that might be relevant to speed up the procedure, like non-specific binding and sliding of the protein on the polymer.

3 Static Properties of Polymers

Since the scope of this dissertation is to study the search dynamics of proteins moving on the surface of a volume for a specific binding site on an enclosed polymer, the final two parts of this chapter focus on a short introduction to polymer physics, which can be found in standard textbooks [88, 89, 90]. After covering some of the most relevant static properties of polymers, an introduction to polymer dynamics will be given.

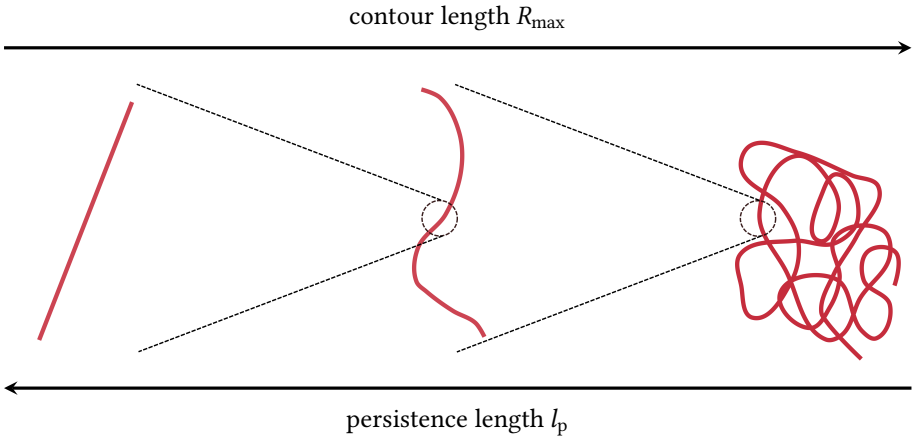
Polymers are macromolecules composed of many repeat units, called monomers. The term monomer refers to single atoms, but also to groups of atoms or even larger structures, as demonstrated in fig. I.6. DNA for example can be described as a linear heteropolymer with four distinct monomers, given by the four nucleotides. Depending on the degree of polymerisation polymers can be very large, highly complex structures. While the chemical properties of a polymer are strongly dependent on every microscopic detail, so that a slight variation in the bond length or angles can for example lead to drastically different features, these details are irrelevant for its physical properties and different polymers exhibit universal behaviour.

On such a scale, where the microscopic properties might be neglected, polymers can be described as linear chains. They exhibit different shapes, from linear, ring shaped and star shaped polymers to bottlebrush structures and polymer networks. We restrict our discussion to linear homopolymers, i.e. polymers composed of a single type of monomer, in contrast to a copolymer or a heteropolymer.

The description of a real polymer by a physical model is not unique. As mentioned above, DNA can be represented by a linear heteropolymer composed of four different nucleotides. However, even bacterial chromosomes contain millions of base pairs and describing every single nucleotide is likely to exceed the desired amount of detail. A coarser description of DNA can therefore be achieved by grouping multiple nucleotides together to a monomer or segment of arbitrary length b, as shown in fig. I.6c. How large this segment length is chosen again depends on the desired amount of detail and significantly determines the appropriate physical model by which to describe the polymer.

On very small scales, where a segment b equals the covalent bonds between atoms, the rigidity of the covalent bonds leads the polymer to appear rather inflexible. On larger scales however, where a segment corresponds to summing up a large number of covalent bonds with different bond angles, the same polymer exhibits some degree of flexibility. As demonstrated in fig. I.7, the same polymer can therefore be described as a rigid rod or as a flexible chain, depending on the resolution of the model. To quantify the appropriate description of a polymer, a measure of its bending rigidity is required.

We define a curvilinear coordinate l which goes from one polymer end $l = 0$ to the contour length $l = R_{\max}$, and tangent unit vectors $\boldsymbol{e}(l)$ describing the changing orientation of the chain. The orientational correlation function is given by $K_{\text{or}}(\Delta l) = \langle \boldsymbol{e}(l)\boldsymbol{e}(l + \Delta l)\rangle$, where $\langle\cdot\rangle$ denotes the ensemble average over all chain conformations. Due to the flexibility of the chain,



(a) Rigid
$R_{max} \ll l_p$

(b) Semiflexible
$R_{max} \approx l_p$

(c) Flexible
$R_{max} \gg l_p$

Figure I.7: Bending rigidity. Bending rigidity of a polymer depending on the contour length R_{max} and the persistence length l_p. (**a**) For very short chains compared to the persistence length thermal fluctuations become negligible and the polymer can be described by a rigid rod. (**b**) When the contour length is of the same order of magnitude as the persistence length, thermal fluctuations become relevant and the appropriate description of the polymer is a semiflexible chain. (**c**) In the limit of very long chains compared to the persistence length the thermal energy dominates and any polymer behaves like a flexible chain.

the orientational correlation vanishes for large distances Δl, so that a measure for the chain stiffness is given by the integral width of K_{or}, which is called the persistence length l_p [91]

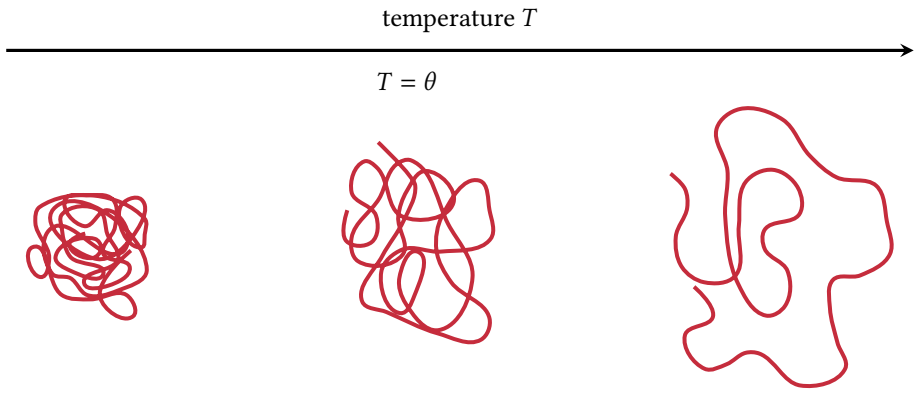
$$l_p = \int_0^\infty K_{or}(\Delta l)\mathrm{d}(\Delta l)\,.$$

The persistence length is therefore the correlation length for orientational fluctuations, such that polymer segments shorter than the persistence length essentially appear as a rigid rod, while polymer segments much longer than the persistence length are orientationally uncorrelated and appear as a flexible chain. When a polymer is divided into segments of length b, the bending rigidity between segments has to be considered for $b < l_p$, whereas for $b > l_p$ the segments are orientationally uncorrelated. In terms of the bending rigidity κ, temperature T and spatial dimension d the persistence length is also given by [92]

$$l_p = \frac{2\kappa}{(d-1)T}\,.$$

With increasing temperature the thermal energy exceeds the bending energy and the polymer becomes more flexible.

Besides the bending rigidity of a polymer, the impact of excluded volume interactions plays a role to determine an appropriate physical polymer model. Excluded volume interactions are highly dependent on the type and the temperature of the surrounding solvent, as demonstrated



(a) Globule
$\nu < 0$

(b) Ideal chain
$\nu = 0$

(c) Random coil
$\nu > 0$

Figure I.8: Excluded volume effects. Excluded volume effects depending on the type and temperature of the surrounding solvent. (**a**) At low temperature (poor solvent) long-range attractive forces between monomers dominate and the polymer adopts a densely packed, globular state. (**b**) At intermediate temperature ($T = \theta$) attractive and repulsive forces balance out and the polymer behaves like an ideal chain. (**c**) At high temperature (good solvent) short-range repulsive forces between monomers dominate and the polymer adopts a swollen coil structure.

in fig. I.8. The different regimes arise due to a competition between enthalpic and entropic effects. At low temperature or in a poor solvent interactions between monomers and solvent molecules are repulsive, leading the solvent to be excluded from the polymer coil and therefore to an effective attraction between monomers. The polymer therefore adopts a densely packed globular state, shown in fig. I.8a. At high temperature or in a good solvent however monomers and solvent mix due to mixing entropy effects, yielding an effective repulsion between monomers. The polymer is in a swollen coil state, as depicted in fig. I.8c. This can be described by the excluded volume parameter ν, which takes into account the effective interaction potential between monomers due to monomer-monomer interactions, monomer-solvent interactions and entropic effects. It is defined as $\nu = 1/2 \int (1 - \exp[U_{\text{eff}}(r)/(k_{\text{B}}T)])\mathrm{d}r$, with an effective interaction potential between monomers $U_{\text{eff}}(r)$ at position r. The interaction potential accounts for the short-range repulsion due to the Pauli exclusion principle and for long-range attraction due to electric dipoles (van der Waals interaction). One of the model potentials to mimic this behaviour is the Lennard-Jones potential, which takes the form

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

with monomer distance r, interaction energy ϵ and monomer diameter σ. The excluded volume parameter is negative when the effective interaction between monomers is attractive and positive when they are repulsive. It therefore defines the spatial volume where a monomer excludes other monomers. At intermediate temperature, the so-called θ-temperature, attractive and repulsive forces cancel each other and the polymer can be well described by an ideal

chain (phantom chain) without excluded volume interactions between monomers ($U_{\text{eff}}(r) = 0$) [88], shown in fig. I.8b.

Overall, the same polymer can be described by a multitude of models, depending on the desired temporal and spatial resolution, the solvent properties and the relation between stiffness and length of the chain. We review some polymer models that are the most relevant to this thesis to find a suitable conformational and dynamical model for DNA.

3.1 Freely Jointed Chain Model

As shown above, when a polymer is partitioned into segments of length $b > l_{\text{p}}$, neighbouring segments are orientationally uncorrelated. One of the simplest descriptions of such a polymer is the freely jointed chain (FJC), also known as the random flight model, which was initially proposed by Kuhn in 1934 [93].

The polymer is described as a chain of $N + 1$ beads connected by N bonds of fixed length b with independent orientation of neighbouring segments. We denote the position vectors of the beads as $\boldsymbol{r} = (\boldsymbol{r}_0, \ldots, \boldsymbol{r}_N)$ and the bond vectors as $\boldsymbol{R} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N)$, with $\boldsymbol{R}_n = \boldsymbol{r}_n - \boldsymbol{r}_{n-1}$ for $n = 1, \ldots, N$. To characterise the size of a polymer, the end-to-end vector of the chain is defined as [94]

$$\boldsymbol{R}_{\text{ee}} = \boldsymbol{r}_N - \boldsymbol{r}_0 = \sum_{n=1}^{N} \boldsymbol{R}_n \,.$$

Due to the independent orientations of the segments the average of the end-to-end vector vanishes ($\langle \boldsymbol{R}_{\text{ee}} \rangle = 0$). The first non-vanishing moment is the mean of the squared end-to-end distance, given by [89]

$$\begin{aligned} \langle \boldsymbol{R}_{\text{ee}}^2 \rangle &= \left\langle \left(\sum_{i=1}^{N} \boldsymbol{R}_i \right)^2 \right\rangle \\ &= \left\langle \sum_{i,j=1}^{N} \boldsymbol{R}_i \boldsymbol{R}_j \right\rangle \\ &= N b^2 \,, \end{aligned} \tag{I.3}$$

where we used $\langle \boldsymbol{R}_i \boldsymbol{R}_j \rangle = \delta_{ij} b^2$, again due to the independently oriented segments. When the segment length is not fixed, but the average segment length is given by $\langle |\boldsymbol{R}| \rangle = b$, the shown relation for the average end-to-end distance is still valid.

Another measure for the size of a polymer is the radius of gyration R_{g}. It is defined as the mean squared distance between monomers and the centre of mass of the polymer $\boldsymbol{r}_{\text{cm}}$

$$\begin{aligned} R_{\text{g}}^2 &= \frac{1}{N+1} \sum_{i=0}^{N} (\boldsymbol{r}_i - \boldsymbol{r}_{\text{cm}})^2 \\ &= \frac{1}{(N+1)^2} \sum_{i,j=0}^{N} (\boldsymbol{r}_i - \boldsymbol{r}_j)^2 \,, \end{aligned}$$

where in the second step the definition of the centre of mass $\boldsymbol{r}_{\mathrm{cm}} = 1/(N+1)\sum_{i=0}^{N}\boldsymbol{r}_i$ was inserted. For ideal chains like the FJC and $N \gg 1$ the mean squared radius of gyration is given by [88]

$$\langle R_{\mathrm{g}}^2\rangle \approx \frac{Nb^2}{6}\,.$$

A real polymer with contour length $R_{\mathrm{max}} = Nb$ and end-to-end distance R_{ee} can be mapped to an equivalent FJC with bond length

$$b = \frac{\langle \boldsymbol{R}_{\mathrm{ee}}^2\rangle}{R_{\mathrm{max}}} =: l_{\mathrm{K}}\,.$$

This defines the so-called Kuhn length l_{K}, which characterises the stiffness of the chain and can be related to the persistence length by $l_{\mathrm{K}} = 2l_{\mathrm{p}}$.

3.2 Simple Random Walk

The FJC model can be further simplified by discretising the polymer segments to lie on a periodic lattice. This can be useful to derive some general properties of this polymer model. We therefore describe the polymer as a random walk of N statistically independent and unbiased steps of length b on a periodic lattice. The microcanonical entropy of such a chain with end-to-end vector $\boldsymbol{R}_{\mathrm{ee}}$ is related to the number $\Omega(N, \boldsymbol{R}_{\mathrm{ee}})$ of distinct random walks with N segments and end-to-end vector $\boldsymbol{R}_{\mathrm{ee}}$ and given by [89]

$$S(\boldsymbol{R}_{\mathrm{ee}}) = k_{\mathrm{B}}\log\left[\Omega(N, \boldsymbol{R}_{\mathrm{ee}})\right], \tag{I.4}$$

with k_{B} being the Boltzmann constant.

Now we consider a one-dimensional random walk of length N with m steps to the right, that ends at position x. There are $\binom{N}{m}$ ways to choose those m steps and hence $\Omega(N,m) = \binom{N}{m} = N!/(m!(N-m)!)$ different chain arrangements of length N with the same end-to-end distance x. The total number of walks with N steps is 2^N and therefore the probability to have an end-to-end distance of x for a walk of length N is $\mathcal{P}(x) = \Omega(N,m)/2^N$. For $x \ll N$ this binomial distribution can be approximated by a Gaussian, as demonstrated in [90]. This leads to

$$\mathcal{P}(x) \approx \frac{1}{2^N}\frac{N!}{(N/2)!(N/2)!}\exp\left[-\frac{x^2}{2N}\right].$$

Using the Stirling approximation for long chains $N! \approx \sqrt{2\pi N}(N/e)^N$ to approximate the prefactors yields

$$\mathcal{P}(x) \approx \sqrt{\frac{2}{\pi N}}\exp\left[-\frac{x^2}{2N}\right].$$

Since $\Omega(N,m)$ is only nonzero for either even or odd x, depending whether N is even or odd, the probability distribution function of x differs by a factor of 2 and reads

$$\Psi(x) = \sqrt{\frac{1}{2\pi\langle x^2\rangle}}\exp\left[-\frac{x^2}{2\langle x^2\rangle}\right],$$

where we also inserted the mean squared average $\langle x^2 \rangle = \int_0^\infty x^2 \Psi(x) \mathrm{d}x = N$.

Consider again a random walk with step size b in three dimensions. Due to isotropy of the walk the three dimensions are independent and the probability distribution function for the end-to-end distance is simply given by the product of the one-dimensional distribution functions. In three dimensions this yields [90]

$$\begin{aligned} \Psi(\boldsymbol{R}_{\mathrm{ee}}) &= \sqrt{\frac{1}{(2\pi)^3 \langle R_x^2 \rangle \langle R_y^2 \rangle \langle R_z^2 \rangle}} \exp\left[-\frac{R_x^2}{2\langle R_x^2 \rangle}\right] \exp\left[-\frac{R_y^2}{2\langle R_y^2 \rangle}\right] \exp\left[-\frac{R_z^2}{2\langle R_z^2 \rangle}\right] \\ &= \left(\frac{3}{2\pi N b^2}\right)^{\frac{3}{2}} \exp\left[-\frac{3\boldsymbol{R}_{\mathrm{ee}}^2}{2Nb^2}\right], \end{aligned} \tag{I.5}$$

where we inserted $\langle \boldsymbol{R}_{\mathrm{ee}}^2 \rangle = Nb^2$ and $\langle R_x^2 \rangle = \langle R_y^2 \rangle = \langle R_z^2 \rangle = Nb^2/3$, which follows from $\langle \boldsymbol{R}_{\mathrm{ee}}^2 \rangle = \langle R_x^2 \rangle + \langle R_y^2 \rangle + \langle R_z^2 \rangle$.

Using $\Psi(\boldsymbol{R}_{\mathrm{ee}}) = \Omega(N, \boldsymbol{R}_{\mathrm{ee}}) / \int \Omega(N, \boldsymbol{R}_{\mathrm{ee}}) \mathrm{d}\boldsymbol{R}_{\mathrm{ee}}$ the entropy defined in eq. (I.4) can be written as

$$S(\boldsymbol{R}_{\mathrm{ee}}) = -\frac{3k_{\mathrm{B}}\boldsymbol{R}_{\mathrm{ee}}^2}{2Nb^2} + \frac{3}{2}k_{\mathrm{B}} \log\left[\frac{3}{2\pi N b^2}\right] + k_{\mathrm{B}} \log\left[\int \Omega(N, \boldsymbol{R}_{\mathrm{ee}}) \mathrm{d}\boldsymbol{R}_{\mathrm{ee}}\right].$$

Since the last two terms are independent of the end-to-end distance, this can be simplified to

$$S(\boldsymbol{R}_{\mathrm{ee}}) = S(0) - \frac{3k_{\mathrm{B}}\boldsymbol{R}_{\mathrm{ee}}^2}{2Nb^2}.$$

Inserting the entropy into the free energy $F = E - TS$ leads to

$$F(\boldsymbol{R}_{\mathrm{ee}}) = F(0) + \frac{3k_{\mathrm{B}}T\boldsymbol{R}_{\mathrm{ee}}^2}{2Nb^2},$$

where we used that the energy does not depend on the chain configuration for an ideal chain and is therefore independent of the end-to-end vector. The free energy $F(0)$ then corresponds to a polymer with both ends at the same point. The free energy is therefore minimised for the zero end-to-end vector, as this leads to the largest number of chain configurations and the highest entropy.

Despite the energy being independent of the chain conformations, an ideal chain exhibits elastic behaviour due to the entropic effects demonstrated above. In comparison with the energy of a Hookean spring ($U = k_s x^2/2$), we can define the entropic spring constant by

$$k_s = \frac{3k_{\mathrm{B}}T}{Nb^2}. \tag{I.6}$$

Note that since for the Gaussian approximation we required that $|R_{\mathrm{ee}}| \ll R_{\mathrm{max}}$, the linear entropic spring looses validity when the chain is stretched close to its maximum extension. However for small forces, experiments have shown that the FJC model describes the measured force-extension curves quite well [95].

3.3 Gaussian Chain Model

As shown in the previous sections, a real chain can be mapped to an equivalent freely jointed chain with segment length l_K and end-to-end vector that follows a Gaussian distribution. Further coarse-graining the model, the FJC can be divided into groups of segments of total length b large enough so that the segment vectors $\boldsymbol{R}$ are normally distributed [94]

$$\Psi(\boldsymbol{R}) = \left(\frac{3}{2\pi b^2}\right)^{\frac{3}{2}} \exp\left[-\frac{3\boldsymbol{R}^2}{2b^2}\right]. \tag{I.7}$$

To account for the entropic elasticity of the segments, such a Gaussian chain is typically represented by beads connected by springs, as demonstrated in fig. I.6d.

3.4 Persistent Chain Model

So far, increasingly coarse-grained polymer models have been considered that are valid for segments of length $b > l_p$ or even $b \gg l_p$. However, for smaller segments $b \leq l_p$, the bending rigidity between segments cannot be neglected and the previously discussed models loose their validity. This can be relevant for very stiff polymers or when considering short sections of dsDNA with rather high resolution.

A widely used model for semiflexible chains is the persistent chain model, also called the wormlike chain (WLC) model or Kratky-Porod chain [96]. The polymer is described as a continuous chain with position $\boldsymbol{r}(l)$ along the chain and constant contour length R_{max}. The simplest model to allow for elastic bending of a continuous rod is to assume a quadratic bending energy. This leads to the Hamiltonian of the WLC of the form

$$\mathcal{H} = \frac{\kappa}{2} \int_0^{R_{max}} \left(\frac{\partial^2 \boldsymbol{r}(l)}{\partial l^2}\right)^2 \mathrm{d}l\,,$$

with the bending stiffness κ [94]. The bending energy defined above leads to an exponential form of the orientational correlation function [97]

$$K_{or}(\Delta l) = \exp\left[-\frac{\Delta l}{l_p}\right].$$

Using the end-to-end vector $\boldsymbol{R}_{ee} = \int_0^{R_{max}} \boldsymbol{e}(l')\mathrm{d}l'$, the mean squared end-to-end distance can be obtained by [97]

$$\begin{aligned}
\langle \boldsymbol{R}_{ee}^2 \rangle &= \int_{l'=0}^{R_{max}} \int_{l''=0}^{R_{max}} \langle \boldsymbol{e}(l')\boldsymbol{e}(l'')\rangle \mathrm{d}l'\mathrm{d}l'' \\
&= 2\int_{\Delta l=0}^{R_{max}} K_{or}(\Delta l)(R_{max} - \Delta l)\mathrm{d}\Delta l \\
&= 2\int_{\Delta l=0}^{R_{max}} \exp\left[-\frac{\Delta l}{l_p}\right](R_{max} - \Delta l)\mathrm{d}\Delta l \\
&= 2l_p R_{max} - 2l_p^2\left(1 - \exp\left[-\frac{R_{max}}{l_p}\right]\right).
\end{aligned} \tag{I.8}$$

First, consider the limiting case for long chains $R_{\max} \gg l_p$, which yields $\langle \boldsymbol{R}_{ee}^2 \rangle = 2 l_p R_{\max}$. Using $2l_p = l_K$ and $R_{\max} = N l_K$ this corresponds to the well-known result for an ideal FJC with segment length l_K, computed in eq. (I.3). For short chains $R_{\max} \ll l_p$ a Taylor expansion of the exponential function in eq. (I.8) to second order yields $\langle \boldsymbol{R}^2 \rangle \approx R_{\max}^2$, i.e. the chain is fully extended, which is the expected result for a rigid rod. The mean squared end-to-end distance of the WLC therefore demonstrates the transition from the rigid rod description of very short chains to a FJC for very long chains compared to the persistence length.

4 Polymer Dynamics

After the short introduction to polymer physics and discussing the most relevant models describing the static properties, we will now focus on polymer dynamics. Since we are concerned with the motion of polymers in a solvent, which is the natural environment of biopolymers in the cell, we start with a general introduction to the diffusive motion of particles in solution.

4.1 Brownian Motion

As first observed by the botanist Robert Brown in 1828 [98], small colloidal particles in liquids exhibit a diffusive motion due to random collisions with the solvent particles, later called Brownian motion. First quantitative explanations were developed by Einstein in 1905 [99] and Smoluchowski [100], both approaching the problem based on partial differential equations which are equivalent to the Fokker-Planck equation.

Given the concentration of particles $c(x,t)$ at position x and time t, the process of diffusion is described by Fick's first law [101], which in one dimension reads

$$j(x,t) = -D \frac{\partial c(x,t)}{\partial x}\,, \tag{I.9}$$

thus relating the flux $j(x,t)$ to the concentration gradient with the diffusion constant D [102]. Fick's law shows that when the concentration is not uniform, there is a particle flux towards regions of lower concentration. Inserting the continuity equation $\partial c/\partial t = -\partial j/\partial x$ eq. (I.9) becomes Fick's second law, the diffusion equation

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

With an external potential $U(x)$ the diffusion equation is turned into the so-called Smoluchowski equation [102]

$$\frac{\partial c(x,t)}{\partial t} = \frac{\partial}{\partial x} \frac{1}{\zeta} \left(k_B T \frac{\partial c(x,t)}{\partial x} + c(x,t) \frac{\partial U(x)}{\partial x} \right)\,. \tag{I.10}$$

The friction coefficient ζ is related to the diffusion coefficient via the Einstein relation $D = k_B T/\zeta$, an early example of the fluctuation-dissipation theorem [103], relating a dissipative force (friction) to thermal fluctuations.

4.2 Langevin Equation

Brownian motion can alternatively be treated by the Langevin description [104], which historically is a first example of a stochastic differential equation. In the Langevin picture the force acting on the Brownian particle due to the small solvent particles is described by a stochastic force, which fluctuates around its average value. Integrating this random force into Newton's equation of motion then yields the Langevin equation. Aside from the random force due to interactions with the fluid, the force acting on a Brownian particle is composed of a friction force and external forces. For a single Brownian particle in one dimension the time evolution of its position $x(t)$ is given by

$$m\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} = -\zeta\frac{\mathrm{d}x(t)}{\mathrm{d}t} - \frac{\mathrm{d}U(x)}{\mathrm{d}x} + f(t)\,.$$

Here, m is the mass of the particle, ζ is the friction coefficient, $U(x)$ is an external interaction potential and $f(t)$ is the random force due to interactions with the solvent. In the limit of low Reynold's number $\zeta/m \gg 1$, i.e. for very high viscosity of the solvent or very small particles, inertial forces become negligibly small and the Langevin equation reduces to the "overdamped" Langevin equation [88]

$$\zeta\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\frac{\partial U(x)}{\partial x} + f(t)\,. \tag{I.11}$$

It can be shown [102] that the distribution of $x(t)$ defined by eq. (I.11) satisfies the Smoluchowski equation (I.10) if the distribution of the random force $f(t)$ is a Gaussian with the moments

$$\begin{aligned}\langle f(t)\rangle &= 0\,,\\ \langle f(t)f(t')\rangle &= 2\zeta k_\mathrm{B}T\delta(t-t')\,.\end{aligned}$$

The mean value vanishes due to isotropy of the solvent and the collisions are assumed to be independent, such that the time correlation is also zero. The variance of the stochastic force follows from the equipartition theorem. For a single particle diffusing in one dimension eq. (I.11) can be easily solved for $U(x) = 0$ (see appendix A1) to show that the mean squared displacement (MSD) fulfils

$$\begin{aligned}\langle(x(t)-x(0))^2\rangle &= \frac{2k_\mathrm{B}T}{\zeta}t\\ &= 2Dt\,,\end{aligned}$$

where the Einstein relation was used in the second step. For a Brownian particle diffusing in d dimensions the mean squared displacement is $\langle(\boldsymbol{r}(t)-\boldsymbol{r}(0))^2\rangle = 2dDt$ [88].

For multiple particles at position $\boldsymbol{r}_i(t)$ with mass m_i, friction tensor $\boldsymbol{\zeta}_{ij}$ and interaction potential $U(\{\boldsymbol{r}_i(t)\})$ the Langevin equation becomes

$$m_i\frac{\mathrm{d}^2\boldsymbol{r}_i(t)}{\mathrm{d}t^2} = -\sum_i \boldsymbol{\zeta}_{ij}\frac{\mathrm{d}\boldsymbol{r}_i(t)}{\mathrm{d}t} - \boldsymbol{\nabla}_{\boldsymbol{r}_i(t)}U(\{\boldsymbol{r}_i(t)\}) + \boldsymbol{f}_i(t)\,, \tag{I.12}$$

and the overdamped Langevin equation

$$\sum_i \zeta_{ij} \frac{\mathrm{d}\boldsymbol{r}_i(t)}{\mathrm{d}t} = -\boldsymbol{\nabla}_{\boldsymbol{r}_i(t)} U(\{\boldsymbol{r}_i(t)\}) + \boldsymbol{f}_i(t)\,. \tag{I.13}$$

For hydrodynamically noninteracting particles $\zeta_{ij} = \delta_{ij}\zeta$ mean value and correlation function of the random force are given by

$$\begin{aligned} \langle f_{i\alpha}(t)\rangle &= 0\,, \\ \langle f_{i\alpha}(t) f_{j\beta}(t')\rangle &= 2\zeta k_\mathrm{B} T \delta_{ij}\delta_{\alpha\beta}\delta(t-t')\,, \end{aligned}$$

with α and β referring to the spatial components.

4.3 Rouse Model

Based on the Langevin description of a bead-spring polymer model, the first model to successfully describe the dynamics of ideal polymers was developed by Rouse in 1953 [105]. The Rouse model is considered to be valid in a dense melt, where hydrodynamic interactions are screened, provided entanglement effects between chains do not play a role. The polymer is described based on the Gaussian chain model as a chain of $N+1$ beads connected by springs of average length b. The interaction potential between neighbouring beads [102]

$$U(\{\boldsymbol{r}_n(t)\}) = \frac{k_s}{2} \sum_{n=0}^{N} (\boldsymbol{r}_n(t) - \boldsymbol{r}_{n-1}(t))^2\,, \tag{I.14}$$

with the entropic spring constant $k_s = 3k_\mathrm{B}T/b^2$ ensures the Gaussian distribution of the segment vectors derived earlier in this chapter. With the given interaction potential eq. (I.14) and the random force $\boldsymbol{f}_i(t)$ acting on bead i at position $\boldsymbol{r}_i(t)$ the overdamped Langevin equation eq. (I.13) becomes

$$\begin{aligned} \zeta \frac{\mathrm{d}\boldsymbol{r}_n(t)}{\mathrm{d}t} &= -k_\mathrm{s}(2\boldsymbol{r}_n(t) - \boldsymbol{r}_{n+1}(t) - \boldsymbol{r}_{n-1}(t)) + \boldsymbol{f}_n(t) \quad \text{for } n = 1, \ldots, N-1 \\ \zeta \frac{\mathrm{d}\boldsymbol{r}_0(t)}{\mathrm{d}t} &= -k_\mathrm{s}(\boldsymbol{r}_0(t) - \boldsymbol{r}_1(t)) + \boldsymbol{f}_0(t) \\ \zeta \frac{\mathrm{d}\boldsymbol{r}_N(t)}{\mathrm{d}t} &= -k_\mathrm{s}(\boldsymbol{r}_N(t) - \boldsymbol{r}_{N-1}(t)) + \boldsymbol{f}_N(t) \end{aligned}$$

and the random force follows a Gaussian distribution with moments

$$\begin{aligned} \langle \boldsymbol{f}_n(t)\rangle &= 0 \\ \langle f_{n\alpha}(t) f_{m\beta}(t')\rangle &= 2\zeta k_\mathrm{B} T \delta_{nm}\delta_{\alpha\beta}\delta(t-t')\,. \end{aligned}$$

Approximating n by a continuous variable, the Langevin equation describing a Rouse chain becomes [88]

$$\zeta \frac{\partial \boldsymbol{r}_n(t)}{\partial t} = k_\mathrm{s} \frac{\partial^2 \boldsymbol{r}_n(t)}{\partial n^2} + \boldsymbol{f}_n(t)\,,$$

with the boundary conditions $\partial \boldsymbol{r}_n(t)/\partial n = 0$ at $n = 0$ and $n = N$.

One of the major features of the Rouse model is its simplicity, which allows for an analytical solution of both the continuous and the discrete equations (see appendices A2.1 and A2.2). Using Fourier transformation the Rouse model is decomposed into independent normal modes $\boldsymbol{X}_p(t)$ (Rouse modes), which are defined by [94]

$$\boldsymbol{X}_p(t) = \frac{1}{N}\int_0^N \boldsymbol{r}_n(t)\cos\left[\frac{p\pi n}{N}\right]\mathrm{d}n\,,$$

with the inverse transform given by

$$\boldsymbol{r}_n(t) = \boldsymbol{X}_0(t) + 2\sum_{p=1}^{\infty}\boldsymbol{X}_p(t)\cos\left[\frac{p\pi n}{N}\right]\,.$$

The normal modes linearise the continuous Langevin equation of the Rouse model, which can be integrated and takes the form

$$\zeta_p\frac{\partial \boldsymbol{X}_p(t)}{\partial t} = -k_p\boldsymbol{X}_p(t) + \boldsymbol{f}_p(t)\,,$$

with $\zeta_0 = N\zeta$, $\zeta_p = 2N\zeta$ and $k_p = 2\pi^2 k_\mathrm{s} p^2/N$ and the random forces $\boldsymbol{f}_p(t)$ which satisfy $\langle f_{p\alpha}(t)\rangle = 0$ and $\langle f_{p\alpha}(t)f_{q\beta}(t')\rangle = 2\zeta_p k_\mathrm{B}T\delta_{pq}\delta_{\alpha\beta}\delta(t-t')$.

The Rouse modes decompose the dynamics of the polymer into independent modes with a certain relaxation rate τ. For all but the zeroth Rouse mode, the correlation function decays exponentially, taking the form [106]

$$\langle X_{p\alpha}X_{q\beta}\rangle = \frac{k_\mathrm{B}T}{k_p}\delta_{pq}\delta_{\alpha\beta}\exp\left[-\frac{t}{\tau}\right]\,.$$

The first mode $p = 1$ has the slowest relaxation time, called the Rouse time τ_R and describes the elongation of the polymer chain. The Rouse time is given by

$$\tau_\mathrm{R} = \frac{(N+1)^2\zeta}{\pi^2 k_\mathrm{s}} = \frac{(N+1)^2 b^2\zeta}{3\pi^2 k_\mathrm{B}T}\,, \tag{I.15}$$

and complies with the time it takes the polymer to diffuse a distance of the order of its own size. The mode with the shortest relaxation time τ_N on the other hand is identified with the relaxation of a monomer. In general, the Rouse mode $\boldsymbol{X}_p(t)$ describes vibrations of a wavelength corresponding to a subchain of N/p segments. The translational motion of the centre of mass is therefore described by the zeroth Rouse mode ($p = 0$) and we identify $\boldsymbol{X}_0(t) = 1/N\int_0^N \boldsymbol{r}_n(t)\mathrm{d}n = \boldsymbol{R}_\mathrm{cm}(t)$ [90].

The dynamics of a polymer within the Rouse model are most commonly characterised by computing the motion of the centre of mass and of single beads respectively. The mean squared displacement of the centre of mass can be easily derived to follow (see appendix A3)

$$g_3 := \langle|\boldsymbol{r}_\mathrm{cm}(t) - \boldsymbol{r}_\mathrm{cm}(0)|^2\rangle = \frac{6k_\mathrm{B}T}{(N+1)\zeta}t = 6D_\mathrm{cm}t\,, \tag{I.16}$$

where we have used the definition of the diffusion constant of the centre of mass $D_{\mathrm{cm}} = \lim_{t\to\infty} 1/(6t)\langle|\boldsymbol{r}_{\mathrm{cm}}(t)-\boldsymbol{r}_{\mathrm{cm}}(0)|^2\rangle$ to obtain $D_{\mathrm{cm}} = k_{\mathrm{B}}T/((N+1)\zeta) = D/(N+1)$. The centre of mass therefore behaves like a Brownian particle, moving diffusively with a reduced diffusion constant $D_{\mathrm{cm}} = D/(N+1)$. The mean squared displacement of a typical segment is [107]

$$\begin{aligned}\tilde{g}_1(t) &= \frac{1}{N+1}\sum_{n=0}^{N}\left\langle(\boldsymbol{r}_n(t)-\boldsymbol{r}_n(0))^2\right\rangle \\ &= 6D_{\mathrm{cm}}t + 4\sum_{p=1}^{N}\langle \boldsymbol{X}_p^2\rangle\left(1-\exp\left[-\frac{t}{\tau_p}\right]\right) \\ &\approx 6D_{\mathrm{cm}}t + \left(\frac{12k_{\mathrm{B}}Tb^2}{\pi\zeta}\right)^{\frac{1}{2}}\sqrt{t}\,,\end{aligned}$$

where the approximation in the last step is valid for $N \gg 1$. At small times up to the Rouse time viscoelastic modes lead to subdiffusion of the monomers with the mean squared displacement scaling as $\propto \sqrt{t}$. For times larger than the Rouse time however the polymer moves as a whole and the mean squared displacement shows normal diffusive behaviour with a scaling $\propto t$. For the mean squared displacement of a single bead we choose the central monomer to avoid unwanted deviations at the edges of the chain. The two different scaling regimes are therefore

$$g_1(t) := \langle|\boldsymbol{r}_{N/2}(t)-\boldsymbol{r}_{N/2}(0)|^2\rangle \propto \begin{cases} t^{0.5} \text{ if } t \ll \tau_{\mathrm{R}} \\ t \text{ if } t \gg \tau_{\mathrm{R}}\end{cases}\,. \tag{I.17}$$

This scaling behaviour will be used in the next chapter to test whether our polymer dynamics simulations agree with the Rouse model.

4.4 Zimm Model

While the Rouse model provides a good approximation in polymer melts, in dilute solutions the force acting on a certain particle causes fluid motion around it, influencing the velocity of other particles [90]. Introducing these hydrodynamic interactions into the Rouse model leads to the Zimm model [108], as the nondiagonal elements in the mobility matrix $\boldsymbol{\mu}_{ij}$ become nonzero. The overdamped Langevin equation becomes [109]

$$\frac{\partial \boldsymbol{r}_i(t)}{\partial t} = \sum_j \boldsymbol{\mu}_{ij}\left(-\boldsymbol{\nabla}_{\boldsymbol{r}_i}U(\{\boldsymbol{r}_i\}) + \boldsymbol{f}_j(t)\right)\,,$$

with the mobility tensor given by the nonlinear Oseen tensor

$$\begin{aligned}\boldsymbol{\mu}_{ii} &= \boldsymbol{I}/\zeta \\ \boldsymbol{\mu}_{ij} &= \frac{1}{6\pi\eta_s|\boldsymbol{R}_{ij}(t)|}[\hat{\boldsymbol{R}}_{ij}(t)\hat{\boldsymbol{R}}_{ij}(t) + \boldsymbol{I}] \quad \text{for } i \neq j\,,\end{aligned}$$

with $\boldsymbol{R}_{ij}(t) = \boldsymbol{r}_i(t) - \boldsymbol{r}_j(t)$ and $\hat{\boldsymbol{R}}_{ij}(t)$ the unit vector. The centre of mass diffusion constant becomes $D_{\mathrm{cm}} \propto k_{\mathrm{B}}TN^{-\nu}/(\eta b)$ with the Flory exponent $\nu = 1/2$ for an ideal chain and $\nu = 3/5$ for an excluded volume chain. The slowest relaxation time, the Zimm time is given by [88]

$$\tau_{\mathrm{Z}} \propto \frac{\eta b^3}{k_{\mathrm{B}}T} N^{3\nu},$$

with the viscosity of the solvent η. Introducing hydrodynamic interactions into the Rouse model therefore leads the diffusion to become faster as the slowest relaxation time becomes shorter.

Design and Testing of a Kinetic Monte Carlo Model

II

In order to study the encounter dynamics of surface proteins with a binding site on an enclosed polymer, a cubic lattice model is established in this chapter, that allows the numerical simulation of the respective search process.

1 Introduction

Despite their novelty, computer simulations have evolved into being a widely recognised complementary approach to research, that bridges the gap between theory and experiment. While the computationally feasible spatial and temporal scale was highly restricted in early computer experiments, advances in computer software, development of multiscale modelling and powerful computers have led to enormous improvements.

After an early paper on Monte Carlo simulations [110], the first computer simulations were reported in 1953 [111], introducing the Metropolis Monte Carlo method to simulate a liquid represented by two-dimensional rigid spheres. Shortly afterwards, molecular dynamics simulations were proposed for the first time and applied to the assembly of hard spheres with an attractive square well potential [112]. Further advancements led to the simulation of a diatomic liquid [113], and water molecules [114, 115], to polymers [116] and other biomolecules [117].

Molecular dynamics (MD) describes a physical simulation method, where particles are moved according to Newton's equations of motion. The interaction potentials between particles are implemented in form of a functional force field, that may be developed from mechanical calculations or empirical knowledge. After the first application of MD simulations to the dynamics of a small globular protein in 1977 [118], improvements in the simulation software, the accuracy of biomolecular force fields and computational power enabled the simulation of increasingly large and complex biomolecules, like the aggregation of dodecylphosphocholine surfactant molecules into a micelle [119]. The use of modern supercomputers extends the time scale accessible to atomistic molecular dynamics simulations to the millisecond scale [120], enabling the direct simulation of fast folding proteins of up to 80 amino acid residues [121].

Despite the increasingly large range of applications for MD simulations, depending on the system of interest there are still limitations. Especially regarding the simulation of solutions, the number of solvent molecules usually outnumbers the number of solutes by far, complicating the task to obtain meaningful and statistically significant simulations. This problem has been solved by numerically integrating the Langevin equation instead of

Newton's equation of motion [122, 123], which was later termed Brownian dynamics (BD) simulations [124]. After successfully simulating the dynamics of small molecules [125], BD simulations were soon applied to polymers [126, 127] and even the self-diffusion [128] and cyclisation kinetics [129] of DNA.

Monte Carlo (MC) methods are stochastic simulations to sample the configuration space under specific thermodynamic conditions. Unlike MD and BD, there is no intrinsic temporal scale in MC simulations. However, in the 1960s MC algorithms started being used to evolve systems also dynamically, e.g. to simulate the self-diffusion of vacancies in binary ordered alloys [130]. The first thorough description of such a kinetic Monte Carlo (kMC) simulation is called the n-fold way [131], providing a self-consistent dynamical interpretation of MC simulations. A very similar approach was developed by Gillespie to simulate chemical reactions [132]. Kinetic Monte Carlo simulations have been successfully applied to the dynamics of biomolecules like the folding of nucleic acid hairpins [133], crystal growth [134], protein folding [135, 136] and the dynamics of flexible polymers [137].

From atomistic MD simulations to Brownian dynamics and coarse-grained kMC simulations, all mentioned simulation procedures have been successfully used to simulate DNA of varying resolution. From the first application to DNA [138], MD simulations have helped to understand force-extension curves [139] and the appearance of kinks and bubbles in supercoiled DNA [140]. While in molecular dynamics DNA molecules of not more than 20 bp length are commonly used, BD simulations have effectively simulated longer DNA strands to investigate the dynamics of DNA supercoiling [141], loop formation [129] and transport processes of short DNA segments [85]. In a recent study we successfully used BD simulations of a discretised WLC model to simulate the dynamics of the buckling transition of dsDNA and RNA [142]. On even larger scales kinetic Monte Carlo simulations can be applied. They have been used to investigate DNA hairpin folding [133], large-scale chromosomal dynamics [143] and DNA hybridisation [144].

Our aim is to simulate a long DNA molecule, representing the genome in a bacterial cell or a chromosome in the nucleus of a eukaryotic cell. Since we want to investigate the dynamics of encounter with a protein on the surface of the surrounding volume, the challenge lies not only in the required length scale but also in the time scale of this process. To be able to achieve this within reasonable computation time, we aim at finding a suitable lattice model for the polymer dynamics. To ensure that the model behaves as expected, we verify that it reproduces Rouse dynamics and compare its results to the ones from off-lattice BD simulations for validation. We describe our choice of modelling and implementing the DNA simulations in the following section before discussing the complete lattice model of the target search dynamics at the end of this chapter.

2 DNA Simulations

2.1 Model

As outlined in the introductory chapter, a polymer like DNA can be represented by a multitude of different models, depending on the type of solvent, the polymer density and the desired resolution of the description. To simulate a long DNA molecule for a time span long

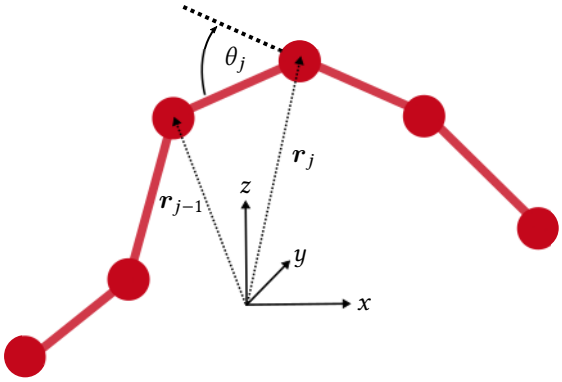


Figure II.1: Discrete wormlike chain. Discrete wormlike chain model of a semiflexible polymer, defining the bending angle θ_j between segments connecting the beads at positions $\boldsymbol{r}_{j-1}$, $\boldsymbol{r}_j$ and $\boldsymbol{r}_{j+1}$.

enough to capture the target search process in the nucleus or bacterial cell is computationally challenging. We are therefore looking for a coarse-grained representation of the DNA. Since bead-spring models that use an entropic spring like the Rouse model loose their validity in strong confinement [145], we choose to model the DNA as a freely jointed chain. Moreover, as will be presented in the following sections, there are multiple widely used lattice representations of a FJC, that allow for an efficient and fast implementation of the polymer dynamics.

When modelling for example the DNA in *E. coli* with around 3.60×10^6 bp inside the cell with volume $\propto 1\,\mu\text{m}^3$ [4] by a FJC with Kuhn segments of length $l_\text{K} = 2l_\text{p} \approx 100\,\text{nm}$ [146] on a cubic lattice, there are 1000 lattice points for 15773 polymer beads. It is therefore natural to model the DNA as a phantom chain, where multiple occupation of a lattice point does not necessarily mean that the polymer has no spatial extension, but it reflects that our choice of lattice is so coarse that many polymer segments can easily fit next to each other on a single lattice site. Moreover, as we intend to simulate the search dynamics in live bacterial cells or eukaryotic nuclei, the presence of topoisomerases that can cut DNA and allow for crossings of the polymer is another argument to favour a phantom chain over excluded volume interactions.

2.2 Brownian Dynamics Simulations

We first define Brownian dynamics simulations of the Rouse model that are used to test the validity of different lattice models for the polymer dynamics. BD simulations make use of the overdamped Langevin equation to advance the system in time. They provide an efficient way of simulating the dynamics of polymers like DNA on length and time scales where inertia does not play a role. The polymers are therefore typically represented by a coarse-grained structure like a bead-spring model. When a spring corresponds to a segment shorter than the persistence length, the polymer can be modelled as a discretised WLC as in our previous study [142]. A bead-spring representation can also be used to model a FJC by using very stiff

springs [147] with an equilibrium length equal to the Kuhn length. On an even coarser scale a spring can correspond to a large number of Kuhn segments that exhibit Gaussian statistics and would therefore be modelled as an entropic spring.

The overdamped Langevin equation (I.13) for hydrodynamically non-interacting particles ($\boldsymbol{\zeta}_{ij} = \zeta\delta_{ij}$) and the bead mobility $\mu_b = 1/\zeta$ can be written as

$$\frac{\mathrm{d}\boldsymbol{r}_i(t)}{\mathrm{d}t} = -\mu_b \boldsymbol{\nabla}_{\boldsymbol{r}_i} U(\{\boldsymbol{r}_i(t)\}) + \mu_b \boldsymbol{f}_i(t)\,,$$

with the random force $\boldsymbol{f}_i(t)$ as defined before. Using the forward (explicit) Euler scheme [148] the discrete time Langevin equation becomes [149]

$$\boldsymbol{r}_i(t+\Delta t) = \boldsymbol{r}_i(t) - \mu_b \Delta t \boldsymbol{\nabla}_{\boldsymbol{r}_i} U(\{\boldsymbol{r}_i(t)\}) + \sqrt{2\mu_b k_\mathrm{B} T \Delta t}\boldsymbol{\eta}_i(t)\,, \tag{II.1}$$

which provides a way to forward the system in time in the BD simulations. We have defined $\mu_b \boldsymbol{f}_i(t) = \sqrt{2\mu_b k_\mathrm{B} T \Delta t}\boldsymbol{\eta}_i(t)$, such that $\langle \boldsymbol{\eta}_i(t)\rangle = 0$ and $\langle \boldsymbol{\eta}_i(t)\boldsymbol{\eta}_j(t')\rangle = 3\delta_{ij}\delta_{tt'}$.

2.2.1 Interaction potentials

The polymer is modelled as an ideal chain of N beads connected by springs with a harmonic potential between nearest neighbours of the form

$$U_\mathrm{s}(\{\boldsymbol{r}_i(t)\}) = \frac{k_\mathrm{s}}{2}\sum_i (|\boldsymbol{r}_i(t) - \boldsymbol{r}_{i+1}(t)| - l_0)^2\,, \tag{II.2}$$

with a spring constant k_s and equilibrium distance between neighbouring beads l_0. This Gaussian chain model can be extended to the discrete wormlike chain by adding a second potential for bending. Assuming a harmonic bending potential it follows [150]

$$U_\mathrm{b}(\{\theta_i\}) = \frac{k_\mathrm{B} T l_\mathrm{p}}{l_0}\sum_i (1 - \cos\theta_i)\,,$$

with the persistence length l_p and the local bending angle θ_i as defined in fig. II.1. The total potential is obtained by $U = U_\mathrm{s} + U_\mathrm{b}$.

In order to validate the dynamics of the lattice polymer models demonstrated later on in this chapter, we run BD simulations of a Rouse chain by using an entropic spring constant $k_\mathrm{s} = 3k_\mathrm{B}T/l_\mathrm{K}^2$ with zero equilibrium length ($l_0 = 0$) and setting the bending potential to zero ($U_\mathrm{b} = 0$). The discrete time Langevin equation is made dimensionless by measuring lengths in terms of the Kuhn length $x' = x/l_\mathrm{K}$ and energies in units of $k_\mathrm{B}T$ (see appendix B1), the time scale being given by the diffusional time scale $\Delta t' = \Delta t \mu_b k_\mathrm{B} T/l_\mathrm{K}^2 = \Delta t D/l_\mathrm{K}^2$. This leads to the dimensionless discrete Langevin equation

$$\boldsymbol{r}_i'(t'+\Delta t') = \boldsymbol{r}_i'(t') + \Delta t' \boldsymbol{F}_i'(t') + \sqrt{2\Delta t'}\boldsymbol{\eta}_i(t')\,,$$

with $\boldsymbol{F}_i(t) = -\boldsymbol{\nabla}_{\boldsymbol{r}_i} U(\{\boldsymbol{r}_i(t)\})$ and dimensionless variables marked by a prime.

The time step $\Delta t'$ in BD simulations has to be chosen carefully to find a compromise between accuracy and speed of the simulations. It was chosen such that the polymer segments are

(a) Random walk **(b)** Linear initialisation **(c)** Random walk (ssBFM)

Figure II.2: Initial chain configurations. (**a**) Random walk on the 3d cubic lattice without diagonal bonds. (**b**) Linear initialisation with segment length equal to the lattice constant $b = a$. (**c**) Random walk on the cubic lattice with allowed bond lengths $b \in \{1, \sqrt{2}, \sqrt{3}\}$.

normally distributed, as expected from the Gaussian chain model. Moreover we verified that the simulated polymer dynamics agree well with the Rouse model (see appendix B1). To analyse the polymer dynamics of the lattice models, different initial polymer configurations are used, defined in fig. II.2. For a random initial configuration the chain is initialised as a random walk on the cubic lattice, as shown in fig. II.2a. To test the relaxation of the polymer also from unusual initial configurations, we introduce a linear initial polymer configuration with segment lengths equal to the lattice constant a, depicted in fig. II.2b. Although this configuration is rather unlikely especially for a polymer in confinement, part of the polymer could be in a stretched out configuration in our target search simulations due to binding of polymer segments to the molecule moving on the surface. Both for a polymer initialised as a random walk and for an initially linear polymer configuration the mean squared deviation of the centre of mass and the mean squared deviation of the central polymer bead obtained from BD simulations of the Rouse model exhibit the expected Rouse scaling (see fig. B.3) according to eq. (I.16) and eq. (I.17). The simulation results with initially linear chain configuration are also shown in fig. II.4. Note that for very small times another regime with linear scaling is observed. This corresponds to so small bead vibrations that the constraints due to nearest neighbour interactions do not play a role yet and the bead moves diffusively.

2.3 Kinetic Monte Carlo Simulations

For a system with a discrete range of states, the time evolution of the probability that the system occupies a certain state can be described by a master equation. A master equation is a set of first-order differential equations of the form

$$\frac{\mathrm{d}p(\boldsymbol{x},t)}{\mathrm{d}t} = \sum_{\boldsymbol{x}'} \left[w_{\boldsymbol{x}\boldsymbol{x}'} p(\boldsymbol{x}',t) - w_{\boldsymbol{x}'\boldsymbol{x}} p(\boldsymbol{x},t) \right], \tag{II.3}$$

with $p(\boldsymbol{x},t)$ the probability that the system is in a state $\boldsymbol{x}$ at time t and transition rates $w_{\boldsymbol{x}'\boldsymbol{x}}$ to go from state $\boldsymbol{x}$ to state $\boldsymbol{x}'$ [151]. The probability $p(\boldsymbol{x},t)$ is increased by the flux from all other states into state $\boldsymbol{x}$, described by the first term in the sum and decreased by the transition out of state $\boldsymbol{x}$, determined by the second term. To propagate the system in time, we are interested in the probability distribution function $\Psi(\tau)$ for the time of the first escape from state $\boldsymbol{x}$. We define the survival function $G(\boldsymbol{x},t) = \mathcal{P}\{\tau \geq t | \boldsymbol{x}(0) = \boldsymbol{x}\}$, i.e. the conditional probability

that the system is in state $\boldsymbol{x}$ at time t given that it was in the same state at time t', which we set without loss of generality to $t' = 0$. Since the left sum in the master equation vanishes for the survival function, the time propagation simplifies to

$$\frac{\partial G(\boldsymbol{x}, t)}{\partial t} = -R(\boldsymbol{x}) G(\boldsymbol{x}, t) \,.$$

We have defined the total rate of change $R(\boldsymbol{x}) = \sum_{\boldsymbol{x}' \neq \boldsymbol{x}} w_{\boldsymbol{x}'\boldsymbol{x}}$. For time independent transition rates, i.e. a Markov process, the solution of this differential equation with initial condition $G(\boldsymbol{x}, t) = 1$ is $G(\boldsymbol{x}, t) = \exp[-R(\boldsymbol{x})t]$. The survival probability is in general related to the probability distribution function of the first escape via $\int_0^t \Psi(t')\mathrm{d}t' = 1 - G(\boldsymbol{x}, t)$. Hence the waiting time distribution can be obtained by

$$\Psi(\tau) = \frac{\partial}{\partial t}(1 - G(\boldsymbol{x}, t))|_{t=\tau} = R(\boldsymbol{x}) \exp[-R(\boldsymbol{x})\tau] \,.$$

To propagate the system in time and space and numerically simulate the master equation (II.3) we are interested in the probability density $g(\tau, \boldsymbol{x}'|\boldsymbol{x})$ that the next event will be a transition from state $\boldsymbol{x}$ to state $\boldsymbol{x}'$ and will occur after a waiting time τ. It can be conditioned on the waiting time distribution $\Psi(\tau)$ and the probability $w_{\boldsymbol{x}'\boldsymbol{x}}/R(\boldsymbol{x})$ to move from state $\boldsymbol{x}$ to state $\boldsymbol{x}'$, yielding [151, 152]

$$g(\tau, \boldsymbol{x}'|\boldsymbol{x}) = \frac{w_{\boldsymbol{x}'\boldsymbol{x}}}{R(\boldsymbol{x})} R(\boldsymbol{x}) \exp\left[-R(\boldsymbol{x})\tau\right] . \tag{II.4}$$

To propagate the system between states, different stochastic algorithms have been described, as summarised in [153]. The most instructive method is the so-called first reaction method introduced in [152]. It makes use of the fact that also for each pathway in the system to transition from state $\boldsymbol{x}$ to $\boldsymbol{x}'$ the waiting time is exponentially distributed with $\Psi_{\boldsymbol{x}'\boldsymbol{x}}(t) = w_{\boldsymbol{x}'\boldsymbol{x}} \exp[-w_{\boldsymbol{x}'\boldsymbol{x}} t_{\boldsymbol{x}'}]$. In the first reaction method the reaction times are computed for all pathways to obtain the smallest time $t_{\boldsymbol{x}_j,\min} = \min_{\boldsymbol{x}_j} t_{\boldsymbol{x}_j}$. The clock is then advanced by $t_{\boldsymbol{x}_j,\min}$ and the system is moved to state $\boldsymbol{x}_j$. Since the first reaction method requires the computation of many random numbers per single time step, it does not provide the most efficient implementation of a kinetic Monte Carlo simulation.

The most commonly used method has been introduced by Bortz, Kalos and Lebowitz [131] and is often referred to as the BKL-algorithm or the n-fold way, but was also introduced by Gillespie as the direct method [152]. In this algorithm the distribution in eq. (II.4) is directly sampled. To this end all transition rates in the system have to be computed to choose the next move with probability $w_{\boldsymbol{x}'\boldsymbol{x}}/R(\boldsymbol{x})$. Then a time step is calculated from the waiting time distribution $\Psi(\tau) = R(\boldsymbol{x}) \exp[-R(\boldsymbol{x})\tau]$. A properly distributed time step is obtained by choosing a random number r from the unit interval $(0, 1)$ and computing $t_{\exp} = -\log[r]/R(\boldsymbol{x})$ [154]. The most challenging step in implementing the direct method is to select and update all possible events and transition rates. When global search and update methods are used, the CPU requirements per kMC event increase linearly with the number of processes and the lattice size [155]. However, our approach of storing events and rates in a container internally using a hash table, allows insertion and removal of elements with constant time complexity on average.

(a) Verdier-Stockmayer move set **(b)** Generalised Verdier-Stockmayer move set

Figure II.3: (Generalised) Verdier-Stockmayer move set. (**a**) The Verdier-Stockmayer move set includes a kink of the end beads by 90° into one of the four possible directions preserving the connectivity of the chain (end-move) and a kink-jump, where a bead is flipped into the opposite corner if the two connected segments are perpendicular to each other. (**b**) In the generalised Verdier-Stockmayer model a crankshaft move is added. When four subsequent polymer beads occupy the four corners of a unit length square on the lattice, the two beads in the middle can flip by 90° in either of the two directions.

Another method is the null event algorithm first introduced in [156]. For N particles in the system and a total rate $r_i(\boldsymbol{x})$ of particle i, an upper bound for the total rate of a particle is defined as $r_{\max} \geq \max_i [r_i]$. The system is then advanced in time using

$$g(\tau, \boldsymbol{x}'|\boldsymbol{x}) = \frac{w_{\boldsymbol{x}'\boldsymbol{x}}}{r_i(\boldsymbol{x})} \frac{r_i(\boldsymbol{x})}{r_{\max}} \frac{1}{N} N r_{\max} \exp\left[-N r_{\max} \tau\right] .$$

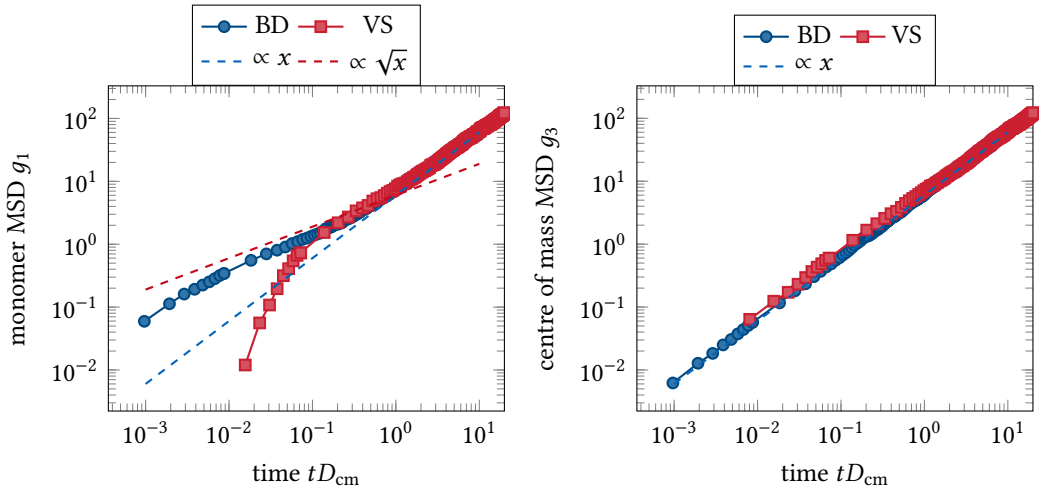
This allows to first select a particle with probability $1/N$, compute its total rate to accept the choice with $r_i(\boldsymbol{x})/r_{\max}$ and then choose a move from $w_{\boldsymbol{x}'\boldsymbol{x}}/r_i(\boldsymbol{x})$. While the previously mentioned methods are rejection-free algorithms, not every event in the null event algorithm is successful. When the total rate of different particles is of the same order of magnitude and does not change considerably during the course of the simulation, the null event algorithm can be advantageous since not all transition rates have to be computed at every time step.

We have implemented both the null event algorithm and the direct method in our simulation framework. As the difference in speed of the two methods depends on the exact configuration of the simulated system and the parameters, we have used both implementations interchangeably. Simulation results shown in this thesis however have been solely obtained using the direct method.

Kinetic Monte Carlo simulations are commonly used with both off-lattice and on-lattice models. Since our aim is to find an appropriate model for the polymer dynamics on a cubic lattice, we will restrict our discussion to some of the most commonly used on-lattice representations.

2.3.1 Verdier-Stockmayer model

The first cubic lattice model to study polymer dynamics using Monte Carlo methods was introduced in 1962 [157] and is nowadays called the Verdier-Stockmayer model. For a cubic lattice with lattice constant a the polymer is represented by a sequence of segments of fixed length a on the grid. The polymer dynamics are implemented by introducing a set of two distinct local moves, as depicted in fig. II.3a. The two end beads can flip by 90° into one



(a) Monomer displacement **(b)** Centre of mass displacement

Figure II.4: Verdier-Stockmayer move set with linear initial configuration. (**a**) The mean squared displacement of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the Verdier-Stockmayer move set (red). The MSD is rescaled using the centre of mass diffusion constant D_{cm}, which is obtained by fitting $g_3(t) = 6D_{\mathrm{cm}}t$. To validate the simulations, the results from BD simulations of the Rouse model are shown in blue. The dashed lines show the scaling of the monomer displacement as expected from the Rouse model, i.e. subdiffusive motion at short times (red) and diffusive motion at long times (blue). (**b**) The corresponding MSD of the centre of mass g_3. The dashed line shows the scaling of the centre of mass displacement as expected from the Rouse model, i.e. diffusive motion with diffusion constant D_{cm}.

of the four possible directions while preserving the connectivity of the chain. In this end-move, if $(\boldsymbol{r}_N - \boldsymbol{r}_{N-1}) \times \boldsymbol{e}_{x_j} \neq 0$ for $j \in \{1, 2, 3\}$ holds, the last bead could be moved to $\boldsymbol{r}'_N = \boldsymbol{r}_{N-1} \pm (\boldsymbol{r}_N - \boldsymbol{r}_{N-1}) \times \boldsymbol{e}_{x_j}$, where we have introduced the unit vectors $\boldsymbol{e}_{x_j}$. The second local move is called a kink-jump, which allows an inner bead to move if the two connected segments are perpendicular to each other. The bead is then moved to the opposite corner $\boldsymbol{r}'_i = \boldsymbol{r}_{i+1} + \boldsymbol{r}_{i-1} - \boldsymbol{r}_i$.

Despite its simplicity the Verdier-Stockmayer model without excluded volume interactions was shown to agree remarkably well with the predictions of the Rouse model [158]. In our attempt to find a lattice model of a FJC that successfully reproduces stable Rouse dynamics, we therefore started with the implementation of the original Verdier-Stockmayer model. When the polymer is randomly initialised, both the mean squared displacement of the centre of mass g_3 and the mean squared displacement of the central bead g_1 show the scaling behaviour that is expected from the Rouse model (see fig. B.4). However, when the polymer is initialised in a linear configuration, strong deviations from the Rouse model arise at short times, shown in fig. II.4. The results from corresponding BD simulations of the Rouse model are shown in comparison, demonstrating that also starting from a linear initial configuration normal Rouse scaling is expected.

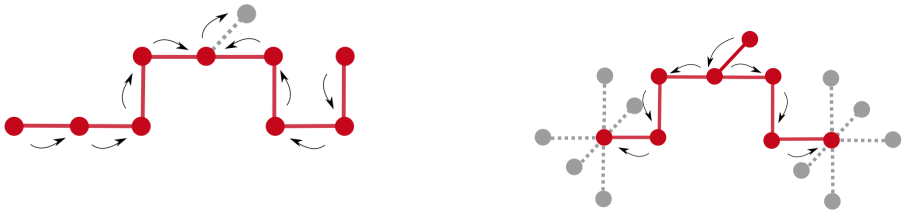
(a) Slithering snake move **(b)** Pivot move

Figure II.5: Global moves. (**a**) In the slithering snake move an end monomer is removed and reattached in a randomly chosen orientation at the opposite chain end, leading to a reptating motion of the polymer. (**b**) In the pivot algorithm a monomer is randomly chosen and either the left or the right portion of the chain is rotated in a random way around this monomer.

A common extension to the Verdier-Stockmayer model to avoid pure reptation-like motion is to add a so-called crankshaft move [159]. As shown in fig. II.3b, if four subsequent polymer beads occupy the four corners of a unit length square on the lattice, the two beads in the middle can flip by 90° in either of the two directions $\boldsymbol{r}_i' = \boldsymbol{r}_{i-1} \pm (\boldsymbol{r}_{i+1} - \boldsymbol{r}_{i-1}) \times (\boldsymbol{r}_i - \boldsymbol{r}_{i-1})$, $\boldsymbol{r}_{i+1}' = \boldsymbol{r}_{i+2} \pm (\boldsymbol{r}_{i+1} - \boldsymbol{r}_{i-1}) \times (\boldsymbol{r}_i - \boldsymbol{r}_{i-1})$. This set of moves is called the generalised Verdier-Stockmayer move set. While starting with a random initial polymer configuration the generalised Verdier-Stockmayer move set still reproduces Rouse dynamics (see fig. B.5), the dynamics of an initially linear chain are still too slow (see fig. B.6). Although adding the crankshaft move allows the polymer to move out of the plane, it does not resolve the problem that for a chain in linear configuration any motion has to diffuse from the polymer ends towards the inner segments.

A possible solution is to add a so-called slithering snake (reptation) move [160]. As depicted in fig. II.5a a polymer bead is removed from one of the chain ends and reattached in a random orientation at the opposite end. When starting with a linear chain configuration the slithering snake move would enable also inner polymer beads to move. This is also reflected in the mean squared displacement of the central monomer when starting from a linear initial configuration (see fig. B.8), which agrees much better with the Rouse model. However, the slithering snake move leads only to Rouse dynamics in the direction in which the chain can reptate, while the dynamics in the other directions are still drastically slowed down (see fig. B.9), which also leads to an unexpected scaling of the centre of mass displacement.

To extend the move options of the polymer further and obtain better agreement with the Rouse model, more local and global moves can be added. In the pivot (wiggle) algorithm [161] depicted in fig. II.5b, a monomer is randomly selected and one of the chain ends is then rotated in a random fashion around this monomer. Another way to improve the mobility of a polymer chain in linear configuration is to define a generalised slithering snake move, as demonstrated in fig. II.6. When the two segments connecting a bead to its neighbours are in a linear configuration, the bead can form a kink into one of the four free directions, while the remaining polymer chain to the right and to the left performs a reptating motion. The reverse move is then to unfold a configuration where the next and the last neighbour of a bead are at the same position by a reptating motion of both sides of the chain and placing



(a) Generalised slithering snake, inwards **(b)** Generalised slithering snake, outwards

Figure II.6: Generalised slithering snake move. (**a**) The inwards move of the generalised slithering snake move is allowed when two segments connecting a bead to its neighbours are in a linear configuration. The bead then forms a kink into a randomly chosen direction and both the left and the right half of the chain perform a reptating motion along their contour. (**b**) The reverse move is the outwards move of the generalised slithering snake move. When the last and the next neighbour of a bead are at the same lattice point, the kink can unfold by a reptating motion of both sides of the chain, placing the end beads in a randomly chosen orientation.

the end beads randomly. Adding the generalised slithering snake move to the generalised Verdier-Stockmayer move set leads to a better agreement of the centre of mass displacement and the monomer displacement with the Rouse model, even when the chain is initially in a linear configuration (see fig. B.10). However, the reptating motion still leads to a fast relaxation in the direction in which the chain is stretched out, while there is too little movement in the perpendicular directions (see fig. B.11).

While adding more moves led to an overall better agreement between the polymer dynamics simulations and the Rouse model, the results were still not satisfactory. In the expectation that relaxing the constraint on the segment length of the chain would yield more realistic polymer dynamics, we therefore turned to a different model, described in the next section.

2.3.2 Single-site Bond-Fluctuation model with Fraenkel spring

Besides the segment length preserving algorithms introduced in the previous section, a widely used model to simulate the dynamics of flexible polymers confined to a lattice is the Bond-Fluctuation model (BFM) [162]. In the original Bond-Fluctuation model a monomer is represented by a unit cube of eight lattice sites, as demonstrated in fig. II.7a. The bond length connecting the monomers is allowed to fluctuate but is restricted to certain bond vectors [163] such that chain crossings are not possible and the model represents a self-avoiding walk (SAW). Monomers are moved by randomly chosen displacements to one of the neighbouring lattice sites. A move is only accepted if it does not result in an overlap of monomers and if the segment vector belongs to the allowed set.

Along with the well-known Bond-Fluctuation model discussed above, Carmesin and Kremer also introduced a so-called single-site Bond-Fluctuation model (ssBFM), which is depicted in fig. II.7b [162]. Here a monomer occupies only a single lattice site and bonds are free to occupy also diagonal configurations with allowed bond lengths $b \in \{1, \sqrt{2}, \sqrt{3}\}$. While the ssBFM is not as frequently used as the original BFM mostly because of ergodicity problems when simulating a SAW, this does not apply to simulating a phantom chain.

Since the allowed bond lengths in the ssBFM are chosen so as to prevent any chain crossings, we can relax this constraint when simulating a phantom chain, as long as the connectivity of the chain is ensured by some other measure that still leads to a valid model of a FJC. In BD simulations a common approach is to approximate a FJC by a bead-spring model. In fact, when Brownian effects are present, a bead-spring model with stiff springs is the more appropriate model compared to a bead-rod model with rigid constraints (Kramers chain), for which the random walk is not an equilibrium configuration, as first pointed out in [164]. A possible way to model a FJC is to use a bead-spring model with stiff Fraenkel springs [147], i.e. Hookean springs with minimum energy at nonzero extension. In the limit of infinitely stiff springs the minimum extension would then correspond to the Kuhn length. Since a stiff Fraenkel spring requires very small time steps in the BD simulation to resolve the fast fluctuations of the spring, a more efficient approach is to use finitely extensible nonlinear elastic Fraenkel (FENE-Fraenkel) springs [145]. Inspired by this common practice in BD simulations, we replace the upper and lower bound of the bond length in the ssBFM by a Fraenkel spring potential

$$U(\{\boldsymbol{r}_n(t)\}) = \sum_{n=1}^{N-1} k_\mathrm{s}(|\boldsymbol{r}_{n+1}(t) - \boldsymbol{r}_n(t)| - l_0)^2\,, \tag{II.5}$$

with spring constant k_s and equilibrium bond length l_0. As in the Bond-Fluctuation model, the polymer moves by displacements of single beads, which can hop to one of the nearest lattice points. The rate of moving a polymer bead then depends on the spring potential as in the Metropolis algorithm [111], i.e. as $k_\mathrm{p} \cdot \min(1, \exp[-U(\{\boldsymbol{r}_n(t)\})/k_\mathrm{B}T])$, where $U(\{\boldsymbol{r}_n(t)\})$ is the configuration-dependent potential energy according to eq. (II.5) and k_p is the simulation parameter for the polymer move rate. To ensure that this model leads to the correct polymer conformations and dynamics, the spring constant is adjusted such that it is stiff enough to represent a FJC and soft enough to reproduce Rouse dynamics (see appendix B3). We chose a spring constant of $k_\mathrm{s} = 10$, which yields a mean segment length of $\langle b \rangle \approx 1.13$ and was found to reliably reproduce Rouse dynamics even when the polymer is initialised in a linear configuration, demonstrated in figs. II.8a and B.17. The slight initial deviation from the BD simulations should not have a major effect on our target search simulations, especially since the Rouse scaling is preserved. This implementation of the polymer dynamics has been used to simulate the target search dynamics of a one-component receptor in *E. coli*, which is described in chapter IV.

2.3.3 Single-site Bond-Fluctuation model

While using a spring potential to ensure the connectivity of the polymer chain leads to the largest possible configuration space while still being restricted to the cubic lattice, other lattice models are far more computationally efficient. In order to speed up the polymer dynamics we therefore use a phantom chain implementation of the original ssBFM with segments of length $b \in \{1, \sqrt{2}, \sqrt{3}\}$. As long as the initial chain configuration is not chosen such that each segment is at its maximum extension (see fig. B.18), the ssBFM of a phantom chain agrees very well with the Rouse model (see figs. B.20 and B.21), shown for a linear initial

(a) Bond-Fluctuation model **(b)** Single-site Bond-Fluctuation model

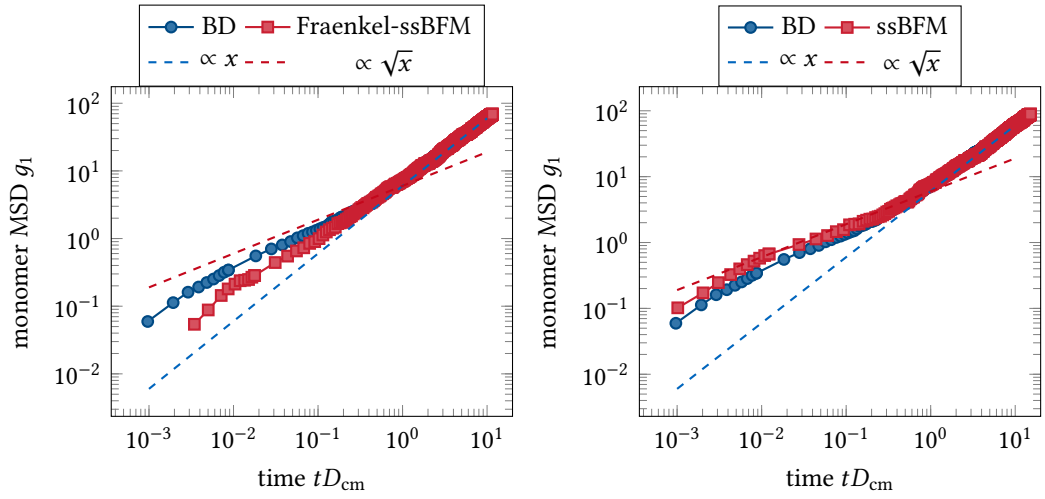
Figure II.7: Bond-Fluctuation models. (**a**) In the Bond-Fluctuation model a monomer occupies a unit cube of eight lattice sites. Monomers are connected by bonds of variable length constrained to a certain bond set to avoid chain crossings. Monomers move by random displacement of a single lattice spacing, ensuring that they do not overlap and the segment length restrictions are fulfilled. (**b**) In the single-site Bond-Fluctuation model a monomer occupies a single lattice site. Monomers are connected by bonds of variable length $b \in \{1, \sqrt{2}, \sqrt{3}\}$, chosen to avoid chain crossings. While the original single-site Bond-Fluctuation model was intended to simulate a SAW, we are simulating a phantom chain and therefore also allow moves as shown on the right, where monomers overlap.

configuration in fig. II.8b. We therefore chose this faster implementation of the polymer dynamics to thoroughly study the encounter dynamics of a confined polymer with a protein at the surface of the confinement, described in chapter III. For this model a random polymer configuration is obtained by a random walk with lattice bond types $b \in \{1, \sqrt{2}, \sqrt{3}\}$, chosen with probabilities $\{3/13, 6/13, 4/13\}$ respectively, as depicted in fig. II.2c [165]. This ensures that given the position of the previously placed polymer bead $\boldsymbol{r}_i$, the position of the next bead $\boldsymbol{r}_{i+1}$ is drawn uniformly from all 13 possibilities. The average squared segment length in the ssBFM is $\langle b^2 \rangle = 2.08$ (see appendix B4).

Alongside with a phantom chain implementation of the ssBFM, where a lattice site is allowed to be occupied by multiple beads and bonds, we also implemented a self-avoiding chain and partial self-avoidance. Self-exclusion of beads and bonds is ensured by forbidding the overlap of beads and bond midpoints, the latter can be shown to lie on a cubic lattice with constant $a/2$ [165]. For partial self-exclusion the overlap of two beads or bonds is penalised by an energy cost ϵ, such that for N_{ov} overlaps the total energy is $E = N_{\mathrm{ov}}\epsilon$ [166]. With the energy difference ΔE between the final and the initial configuration, the polymer move rate is adjusted by the Metropolis rate $\min\left(1, \exp\left[-\Delta E / k_{\mathrm{B}}T\right]\right)$ [111].

3 Target Search Simulations

After establishing a way to simulate a flexible polymer as a FJC on a cubic lattice, a simulation framework was built to numerically simulate the target search of a protein moving on the surface of a volume for a binding site on a polymer inside the volume. Since we are interested in the search times, i.e. long time scales that involve many steps on the grid, the discrete description is a good approximation.



(a) ssBFM with Fraenkel spring

(b) ssBFM

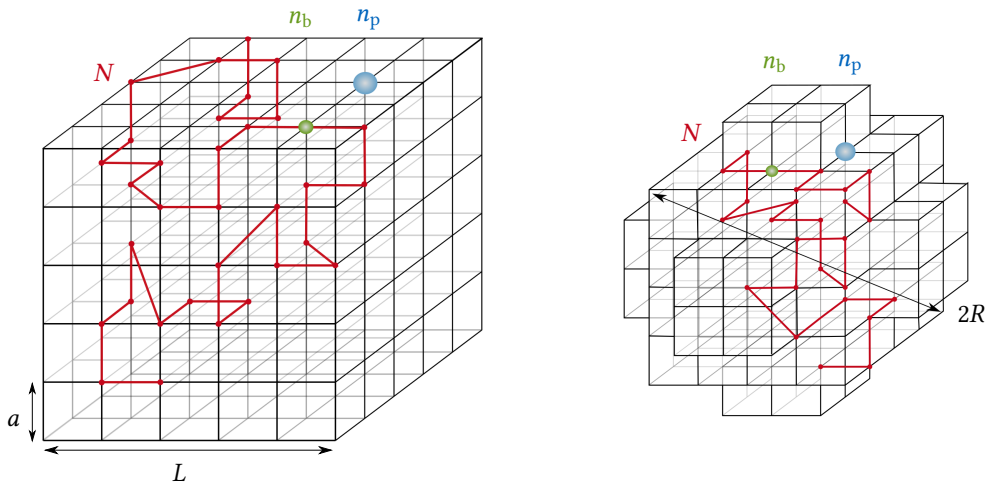
Figure II.8: Rouse scaling of the ssBFM and ssBFM with Fraenkel spring. (**a**) The mean squared displacement of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the ssBFM with Fraenkel spring constant $k_s' = 10$ (red). The MSD is rescaled using the centre of mass diffusion constant D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations, the results from BD simulations of the Rouse model are shown in blue. The dashed lines show the scaling of the monomer displacement as expected from the Rouse model, i.e. subdiffusive motion at short times (red) and diffusive motion at long times (blue). (**b**) The corresponding MSD of the central monomer simulated with a phantom chain implementation of the ssBFM (red) in comparison to results from BD simulations of the Rouse model (blue).

3.1 Configuration Space

Our model is based on a simple cubic lattice with constant a, forming either a cube of length L or the approximation of a sphere of radius R and volume V, depicted in fig. II.9. The polymer is represented by N beads with coordinates $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ placed on the grid points, connected by $N-1$ bonds for which also diagonal conformations are allowed. From all polymer beads, n_b beads are defined as the specific binding sites. Each of the n_p proteins occupies a single lattice point on the surface of the volume. Since the polymer is modelled as a FJC, the mean segment length is chosen to be equal to the Kuhn length l_K. For the ssBFM with Fraenkel spring constant $k_s' = 10$ this yields $\sqrt{\langle b^2 \rangle}a = l_K$ with $\langle b^2 \rangle = 1.32$, for the original ssBFM we obtained $\langle b^2 \rangle = 2.08$ (see appendix B4).

3.2 Dynamics

The polymer dynamics are implemented by random displacements of single monomers with a rate k_p, as described in the previous section, but confined to the volume V. The polymer moves are demonstrated in fig. II.10b, with an example configuration on the right that would be restricted by an energy penalty ϵ for a partially self-avoiding chain or forbidden



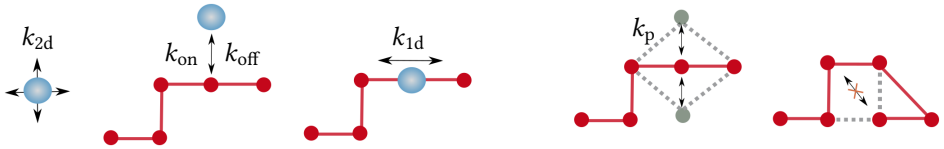
(a) Cubic volume **(b)** "Spherical" volume

Figure II.9: Lattice model. Model for the kinetic Monte Carlo simulations of the encounter dynamics of a confined polymer with a moving protein at the surface of the confining volume. The volume represents a cell, nucleus or vesicle with the polymer (red) of length N inside and n_p proteins (blue) on the surface. The n_b specific binding sites on the polymer are shown in green (**a**) Cubic lattice of length L with lattice constant a. (**b**) Approximation of a sphere of radius R on the cubic lattice.

for full excluded volume interactions of the polymer. The proteins are moving by random displacements to one of the nearest lattice sites on the surface, i.e. performing a 2d random walk. The transition rate for moving a protein is connected to the 2d diffusion coefficient of the protein D_{2d} and the lattice constant a via $k_{2d} = D_{2d}/a^2$. Inspired by the ability of membrane-integrated transcription factors to bind non-specifically to DNA, the proteins are allowed to bind to any polymer bead occupying the same grid point with a rate k_{on} and unbind with k_{off}. Bound proteins move with a rate $k_{1d} = D_{1d}/a^2$ to a neighbouring polymer bead that is at the surface (sliding), while bound polymer beads are restricted to move along the surface with the usual rate k_p, thereby dragging the protein. The different moves of the protein are depicted in fig. II.10a. The rate of moving a polymer bead k_p is adjusted numerically such that the free polymer (on an infinite lattice) displays a centre of mass diffusion coefficient of D_0/N, where D_0 is the diffusion coefficient of a monomer. This is necessary because reduced move opportunities of beads due to excluded volume interactions, a limited move set or segment length restrictions lead in general to a chain length dependence of the bead diffusion constant D_0 [157, 167]. The different move rates are summarised together with other system parameters in Table II.1.

3.3 Simulation Process

The simulation is started with initialising the desired simulation volume and defining an initial configuration of all particles. Most simulations are initialised with a random configuration



(a) Protein movement **(b)** Polymer movement

Figure II.10: Protein and polymer movement. (**a**) Proteins diffuse on the surface with rate k_{2d}, bind unspecifically to the polymer with rate k_{on}, slide along the polymer when bound with rate k_{1d} and unbind again with rate k_{off}. (**b**) The polymer moves by random displacement of single beads with rate k_p, without excluded volume interactions between beads or bonds. For a (partially) self-avoiding chain, the occupancy of a lattice site by multiple beads and the crossing of bonds is forbidden (or reduced by an energy penalty ϵ), an example configuration being demonstrated on the right.

of the polymer inside the volume, n_b beads being defined as specific binding sites and n_p proteins being placed either randomly onto the surface or at a fixed position as defined in a configuration file. While in most of the shown results the simulated polymer is a linear chain, we also implemented the option to simulate a circular polymer.

With the dynamics as described above, the system is then evolved in time until a binding site hits a protein at a search time τ, thus ending the search process. When both the polymer and the proteins are placed randomly, a specific binding site and a protein can be placed at the same lattice site, leading to a measured search time of $\tau = 0$. We perform $\geq 10^3$ such simulations for each parameter set and compute the search time distribution $\Psi(\tau)$ and the mean search time $\langle\tau\rangle$ to analyse the search process.

4 Discussion

With the aim to construct a simple cubic lattice model to efficiently simulate the target search dynamics of a protein moving on the surface of a volume for a binding site on an enclosed polymer, various polymer models have been tested and evaluated. As a measure for the correct dynamics, we compared to the Rouse model and to off-lattice Brownian dynamics simulations of a Rouse chain.

Since bead-spring models using an entropic spring loose their validity in strong confinement [145], the polymer was chosen to be modelled as a FJC. While we have added the option of (partial) self-avoidance of the polymer, both the coarseness of our model and the presence of topoisomerases that can cut DNA and allow for crossings of the chain led us to choose a phantom chain implementation to simulate the polymer.

While the Verdier-Stockmayer move set led to polymer dynamics that were in good agreement with the Rouse model when the polymer was initialised in a random configuration, starting from a linear polymer configuration led to a considerable slow down of the polymer relaxation and deviations from the expected Rouse dynamics. Adding further moves to the set, like the crankshaft move and a slithering-snake move, led to some improvement of the dynamics, however the result was still not satisfactory.

N	number of polymer beads
n_p	number of proteins
n_b	number of binding sites
a	lattice constant
L	grid size in direction x, y, z (cube)
R	radius of spherical grid
k_p	move rate of the polymer
k_2d	2d move rate of the protein
k_1d	sliding rate of the protein
k_on	non-specific binding rate
k_off	non-specific unbinding rate
ϵ	energy penalty for partial self-exclusion
D_0	monomer diffusion constant

Table II.1: System parameters. Parameters defining the dimensions of the simulation volume, the polymer and the proteins, as well as kinetic and binding constants of the target search simulations.

The rigid constraint on the segment length was therefore relaxed and a single-site Bond-Fluctuation model [162] was employed, where also diagonal configurations are allowed for the bonds connecting the polymer beads. Inspired by the common practice in BD simulations to replace the rigid bonds in a FJC by stiff springs, we developed a ssBFM with stiff Fraenkel springs connecting the beads. As in the ssBFM the polymer is moved by random displacement of single beads, the move rate being dependent on the spring potential as in the Metropolis algorithm. When the spring constant is chosen stiff enough to represent a FJC and soft enough to exhibit Rouse dynamics, this implementation yielded a very good agreement with the Rouse model, even when the polymer was initialised in a linear configuration.

While the ssBFM with Fraenkel springs leads to the largest configuration space of the polymer, while still being constrained to the lattice, it does not provide the most efficient implementation of a lattice polymer model. Implementing a phantom chain version of the original ssBFM resolves this problem but with the drawback that there are configurations where the segments are at their maximum extension, leading to slower relaxation dynamics than expected from the Rouse model. However, since also a linear chain configuration with segment length a led to a very good agreement with the Rouse model, we considered the dynamics still stable enough for our purpose.

Using either the ssBFM with Fraenkel springs or the original ssBFM with optional (partial) self-avoidance of the chain we constructed a simulation framework for the target search of a protein on the surface of a volume for a binding site on an enclosed polymer. Besides 2d diffusion of the proteins on the surface we implemented non-specific binding to the polymer as well as 1d sliding along polymer segments that are at the surface. The simulation framework allows a thorough investigation of the respective search process. Varying the different simulation parameters their impact on the dynamics can be studied separately to identify strategies to speed up the procedure. Moreover, adjusting the parameters to a certain

biological process the simulations can be compared to experiments on membrane-integrated transcription factors or mRNA molecules searching for a nuclear pore complex.

Search Dynamics of a Polymer in Confinement

III

Having established a cubic lattice model that allows us to numerically simulate the search dynamics of a protein moving on the surface of a volume for a binding site on a polymer inside the volume, we extensively study this process in order to identify possible pathways to speed up the procedure. The following chapter is adapted from the manuscript that will be published in [168].

1 Introduction

For many cellular functions it is of vital importance that a binding site on a biopolymer and a protein on the surface of a confining domain encounter each other. It has therefore become of increasing interest to biological physics to characterise these search processes.

As discussed in chapter I, the dynamics of a cytoplasmic transcription factor searching for a target on the polymer have been thoroughly studied in the past decades. Also the narrow escape problem, concerning the mean first passage time of a Brownian particle to reach a small absorbing window in the otherwise reflecting boundary of a bounded domain is a well studied problem. While the narrow escape problem is also applicable to small polymers, like mRNA molecules that have to locate nuclear pores and exit from the nucleus in order to be translated [30], it does not apply to longer polymers. However, many biological processes require the colocalisation of a specific binding site on a large polymer with a protein located at the boundary of the confining domain. Examples are DNA double-strand breaks that are relocating to nuclear pore complexes, which has been shown to facilitate the repair process [31] and recruitment of active genes towards NPCs [35]. An example of boundary-located proteins binding to the chromosomal DNA in prokaryotes are membrane-integrated transcription factors like members of the ToxR receptor family.

While the biological examples demonstrate that locating a specific binding site on a long polymer from the surface of a confining volume is feasible, the dynamics of these processes are still highly uncharacterised. In analogy to the facilitated diffusion model for cytoplasmic transcription factors the main question is to which extent elaborate strategies are required in order to facilitate the search process. We employ kinetic Monte Carlo simulations of our coarse-grained cubic lattice model to analyse how a protein moving on the surface of a confining volume locates a binding site on an enclosed polymer.

We find that reflecting boundaries of the confining volume lead the search process to become independent of polymer length in the limit of large polymer densities for ideal chains. An optimal relative binding rate is identified, showing that the search is facilitated by transient

tethering of the protein to the polymer, as well as an optimal sliding rate for permanently bound proteins. We explain the stochastic processes leading to the search time distributions observed depending on the protein diffusion rate, which dominates the search process for intermediate diffusion rates. Moreover, the slowdown of the search due to exclusion of the polymer from the corners of the cubic lattice is discussed.

2 Results

Before discussing the results of our numerical simulations in more detail, a brief overview of the phenomenology that we observe based on the search process is given, as summarised in fig. III.1. In the results presented in this chapter, distance is measured in units of the lattice constant a and time in units a^2/D_0, with the monomer diffusion constant D_0. Unless otherwise stated a linear ideal polymer (phantom chain) with the specific binding site at the centre of the polymer was simulated inside a cubic volume. Moreover, we used a random initialisation of the polymer and the protein.

The dependence of the mean search time $\langle\tau\rangle$ on polymer length N is depicted in fig. III.1a for simulations of a single static protein initially placed at the face centre of the cubic volume. The respective simulations did not include non-specific binding and unbinding and a single binding site was placed at the centre of the polymer. The search time increases with polymer length for short chains but saturates quickly and reaches a plateau as the chain length increases further.

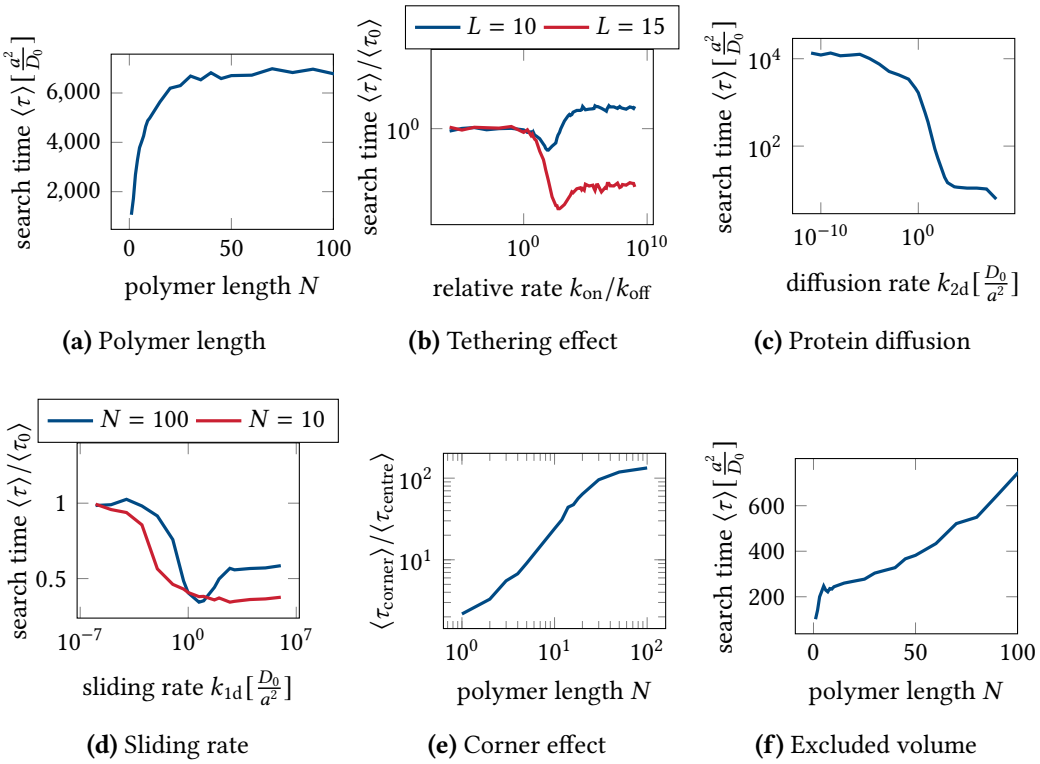
When a single moving protein is allowed to non-specifically bind and unbind to the polymer, an optimal relative binding rate k_{on}/k_{off} is observed, shown in fig. III.1b. For decreasing polymer density (i.e. increasing box size L or decreasing polymer length N) the minimum is shifted towards larger values of k_{on}/k_{off}.

Varying the surface diffusion rate of a single protein that is not allowed to non-specifically bind and unbind to the polymer, the mean search time is a monotonically decreasing sigmoidal function with increasing protein diffusion k_{2d}. As depicted in fig. III.1c, the search time varies between a high plateau for slow protein diffusion and a low plateau for fast protein diffusion with an intermediate regime where the search is largely dominated by the protein diffusion rate.

Varying the sliding rate of a moving protein that is approximately permanently bound, i.e. for high non-specific binding rates and low unbinding rates reveals an optimum at around $k_{1d} \approx 1$, shown in fig. III.1d. The position of the optimum is highly conserved, depending solely on the polymer diffusion rate.

To evaluate the effect of the geometry of the cubic volume on the search time a single static protein is placed either at the corner or at the face centre of the cube. fig. III.1e shows the ratio of the mean search time to reach the protein at the corner or at the face centre respectively, where non-specific binding and unbinding was not included. The ratio increases dramatically with polymer length N until it reaches a plateau.

When excluded volume interactions of the polymer are considered, varying the polymer length in simulations of a single static protein without non-specific binding and unbinding leads to a very different behaviour than for a phantom chain. After an initial strong increase



(a) Polymer length **(b)** Tethering effect **(c)** Protein diffusion

(d) Sliding rate **(e)** Corner effect **(f)** Excluded volume

Figure III.1: Phenomenology of the search process. (**a**) Mean search time $\langle\tau\rangle$ as a function of polymer length N for simulations of a single static protein placed at the face centre of the cubic volume without including non-specific binding and unbinding. A single binding site was placed at the middle of the polymer. (**b**) Mean search time $\langle\tau\rangle$ normalised by the mean search time $\langle\tau_0\rangle$ for $k_{on} = 0$ as a function of k_{on}/k_{off} for different box sizes L. For decreasing polymer density the minimum is shifted towards larger values of the relative binding rate. (**c**) Mean search time $\langle\tau\rangle$ as a function of protein diffusion rate k_{2d} without non-specific binding and unbinding to the polymer. (**d**) Mean search time $\langle\tau\rangle$ normalised by the mean search time $\langle\tau_0\rangle$ for $k_{1d} = 0$ for different polymer lengths and a moving protein that is approximately permanently bound, i.e. with high binding rate and low unbinding rate. (**e**) Ratio of the mean search time to reach the protein at the corner $\langle\tau_{corner}\rangle$ and the mean search time to reach the protein at the face centre of the cube $\langle\tau_{centre}\rangle$ as a function of N. A single static protein was placed either at the corner or at the face centre of the cubic box without including non-specific binding and unbinding. (**f**) Mean search time $\langle\tau\rangle$ as a function of polymer length N for a polymer with excluded volume interactions. A single static protein was placed at the face centre of the cube without including non-specific binding and unbinding.

the mean search time reaches a small plateau but then it diverges as the polymer density in the volume increases, demonstrated in fig. III.1f.

2.1 Polymer Length

To investigate the impact of polymer length and the size of the confining volume on the target search problem we first consider simulations of a single binding site and a single static surface protein without non-specific binding and sliding.

2.1.1 Search time becomes independent of polymer length

Placing the static protein at the face centre of the cubic box we measure the mean search time $\langle\tau\rangle$ as a function of polymer length N for different sizes of the confining volume L. Figure III.2a shows the mean search time normalised by $\langle\tau_1\rangle$, the mean search time of a monomer. The search time first increases with N, before it reaches a plateau and becomes independent of polymer length. As apparent from the plot, both the height of the plateau $\langle\tau_{\text{plateau}}\rangle$ and the crossover polymer length N_{cross}, i.e. the polymer length for which the initial increase transitions into a plateau, depend on the volume of the confinement.

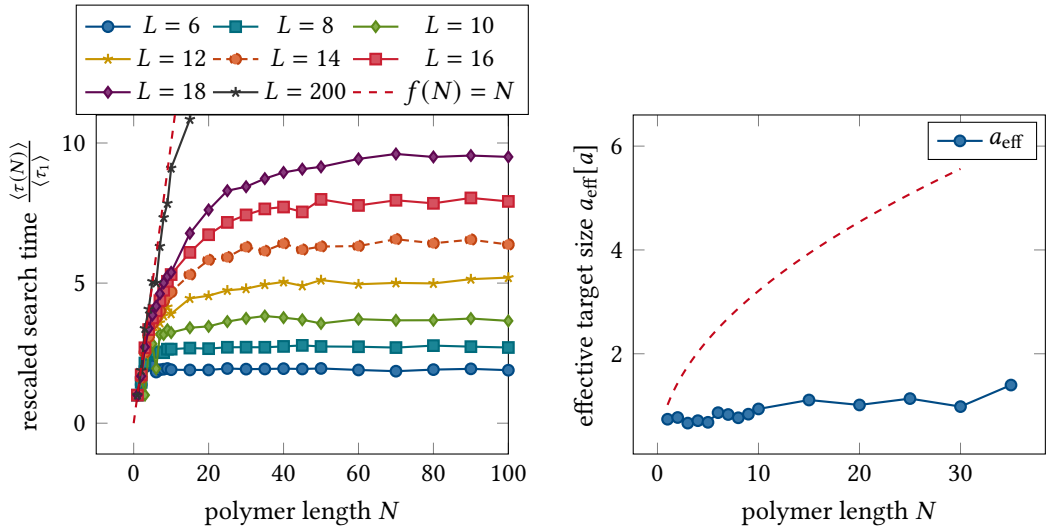
2.1.2 Slow down of linear increase due to compact search

When the polymer is small compared to the confining volume, i.e. for the initial increase of the search time in fig. III.2a, two different regimes can be observed. As stated before, for a Brownian particle, the mean search time (NET) to reach a circular target of radius a_t at the boundary of a volume V is given in first order approximation by Lord Rayleigh's formula [62, 64]

$$\langle\tau\rangle \approx \frac{V}{4\pi D_0 a_t}, \tag{III.1}$$

which agrees well with our simulations (see fig. D.1a). For very small polymers $N \geq 1$ the whole polymer has to move close to the surface to enable the binding site to encounter with the protein and end the search process. Hence the dynamics are largely dominated by the centre of mass dynamics of the polymer, which behave like a particle with a reduced diffusion constant $D_{\text{cm}} = D_0/N$, where D_0 is the monomer diffusion constant. For very small polymers we expect $\langle\tau\rangle \propto VN/(4D_0 a_t)$. Therefore the mean search time in fig. III.2a increases linearly with polymer length for very small polymers, as demonstrated by the red dashed line.

In a recent study [169] the general theory to evaluate the MFPT for regular random walks in bounded domains has been extended to non-Markovian processes and applied to a reactive monomer of a Rouse chain looking for a target in a confined domain. By computing the distribution of the positions of all monomers at the instant of reaction the MFPT is calculated, as well as asymptotic relations for the scaling dependence with the volume. In the large volume limit the scaling relations for different sizes of the target a_t compared to the polymer



(a) Polymer length dependence

(b) Effective protein size

Figure III.2: Polymer length dependence. (**a**) Mean target search time $\langle\tau\rangle$ as a function of polymer length N for different box sizes, normalised with the NET $\langle\tau_1\rangle$. The red dashed line is $\propto N$. The static protein is placed at a the face centre of the cubic volume, the polymer is initialised with a random walk and the mean search time is taken over $\geq 10^3$ runs. After an initially strong increase the mean search time saturates at a certain, box size dependent level. In the limit of small polymer length, polymers behave much like Brownian particles, leading to the initial linear increase. (**b**) Effective protein size a_{eff} for different polymer sizes N obtained by fitting the search time normalised to the polymer length to $\langle\tau\rangle/N = V/(4D_0 a_{\text{eff}})$ (blue). The effective protein size is approximately constant, in agreement with eq. (III.4). The effective protein size according to eq. (III.3) is shown for comparison (red dashed line).

are given by

$$\langle\tau\rangle \propto \begin{cases} \dfrac{V}{4\pi D_0 a_{\text{t}}} & \text{if } a_{\text{t}} \ll \dfrac{l_0}{\sqrt{N}}\,, \qquad \text{(III.2)} \\ \dfrac{V}{4\pi D_{\text{cm}} a_{\text{eff}}} & \text{if } \dfrac{l_0}{\sqrt{N}} \ll a_{\text{t}} \ll l_0\sqrt{N}\,, \qquad \text{(III.3)} \\ \dfrac{V}{4\pi D_{\text{cm}} a_{\text{t}}} & \text{if } a_{\text{t}} \gg l_0\sqrt{N}\,, \qquad \text{(III.4)} \end{cases}$$

where $a_{\text{eff}} \approx 0.71 b\sqrt{N}$ is an effective target radius and l_0 the equilibrium bond length. The three domains reflect the different scaling regimes of the mean squared displacement of a monomer in a Rouse chain. At small length scales $< l_0$ it moves diffusively with diffusion constant D_0, at intermediate length scales between l_0 and the radius of gyration $l_0\sqrt{N}$ it moves subdiffusively and at large scales $> l_0\sqrt{N}$ it again moves diffusively with a reduced diffusion constant $D_{\text{cm}} = D_0/N$. For very small polymers our simulations are in agreement with eq. (III.4), which can be seen by the initial linear increase of the mean search time, because in this case the protein is relatively large compared to the radius of gyration.

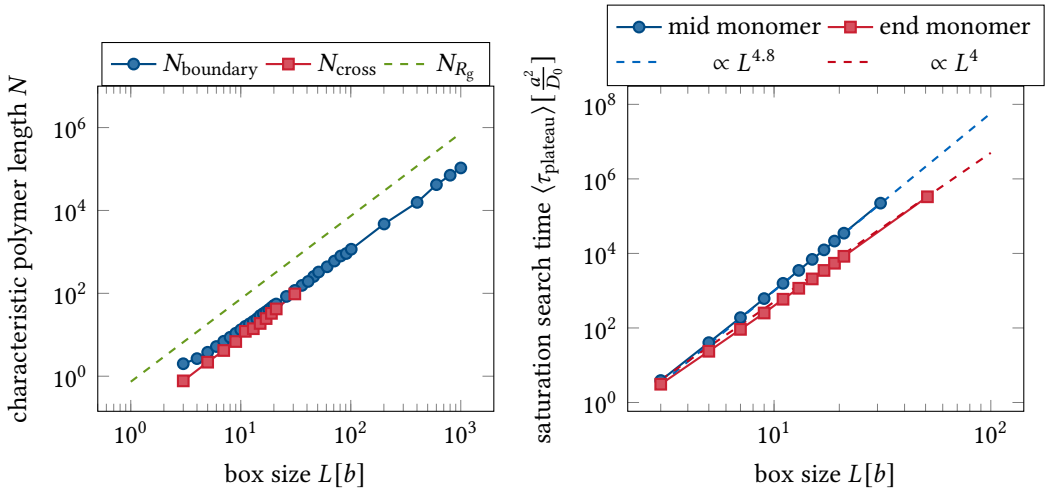
For longer polymers, $N \gg 1$, but still small compared to the confinement, we observe for the mean search time $\langle\tau\rangle \propto \sqrt{N}$, in accordance with eq. (III.3). In this regime, the movement of the binding site is not dominated by the centre of mass diffusion of the polymer anymore. Instead, the subdiffusive motion of a monomer with a mean squared displacement $\langle\Delta\boldsymbol{r}^2\rangle \propto \sqrt{t}$ (in contrast to the diffusive motion of a Brownian particle $\langle\Delta\boldsymbol{r}^2\rangle \propto t$) leads to a compact search [170]. This means, that the searcher site densely explores the space around the protein and hits it on average faster than a Brownian particle with the same diffusion constant. Characteristic for the compact search is that the search time is independent of the actual protein size [169].

We compute the effective protein size a_{eff} by fitting the search time normalised to the polymer length to $\langle\tau\rangle/N = V/(4D_0 a_{\text{eff}})$ (see fig. D.1b and table D.1) in the limit of large volume, leading to an approximately constant value for the effective protein size independent of polymer length, shown in fig. III.2b. This is in accordance with eq. (III.4). When we plot $\langle\tau\rangle/N$ as a function of L for e.g. $N = 20$, we first observe an increase steeper than $\propto L^3$, because the polymer is large compared to the volume, corresponding to the plateau regime in fig. III.2a. The transition to $\propto L^3$ happens when the $\propto \sqrt{N}$ regime in fig. III.2a is reached and persists for the linear regime. Therefore, in order to compute the polymer length dependent effective protein size according to eq. (III.3) would require the fit to an intermediate range of V, which is difficult to define.

2.1.3 Decorrelation of subchains leads to search time plateau

After the increase of the mean target search time with polymer length, the curve flattens and reaches a plateau. In this regime the large volume asymptotics in eqs. (III.2–III.4) is not valid anymore, instead the search time becomes independent of N. The plateau arises due to the confinement of the polymer, which is no longer small compared to the box size.

To quantify this, we define a crossover polymer length N_{cross} when the plateau is reached for different confining volumes, shown in fig. III.3a in red. Since the linear increase in fig. III.2a is very short, we compute the crossover polymer length by fitting the initial increase of the search time with polymer length to $f(N) = c\sqrt{N}$ with a fit parameter c and taking the intersect with the plateau value (see fig. D.3a). We compare the crossover polymer length to the length of a polymer that has a radius of gyration of half the box size $N_{R_g} = 3L^2/(2b^2)$, i.e. a polymer that just fits inside the volume, which is shown by the green dashed line in fig. III.3a. While the scaling ($\propto L^2$) agrees approximately with N_{cross}, the estimate from the radius of gyration is much higher than the computed crossover polymer length. If we take e.g. the curve for $L = 8$ in fig. III.2a and compute the number of beads of a polymer with a radius of gyration of half the box size ($R_g = 4$), this gives $N_{R_g} = 6R_g^2/b^2 \approx 47$ (for the ssBFM $\langle b^2\rangle \approx 2.08$). However, the plot shows that the plateau is already well established for $N < 20$. This observation, that considering the radius of gyration the search time plateau would be expected to be shifted towards longer polymers has been reported before [85]. In this previous study Brownian dynamics simulations of a polymer inside a spherical confinement were used to investigate the time it takes a single monomer to locate a single static protein at the boundary of the confinement. For a polymer with gyration radius $R_g \approx 150$ they obtained a crossover polymer length of $N_{\text{cross}} \approx 10–15$, in good agreement with our simulations. We go



(a) Crossover polymer length

(b) Saturation search time

Figure III.3: Saturation search time and crossover polymer length. (**a**) Crossover polymer length N_{cross} when the plateau is reached as a function of box size. The crossover is computed by fitting the initial increase (from $N = 1$ to L) of the search time with polymer length to $f(N) = c\sqrt{N}$ and taking the intersect with the plateau value as the crossover polymer length. In comparison the length of the polymer with a radius of gyration of half the box size $N_{R_g} = 3L^2/(2b^2)$ is shown by the green dashed line. The blue data show the average segment length for the binding site between two encounters with the boundary. (**b**) Saturation search time as a function of box size for the binding site at the middle of the polymer (blue) and the end of the polymer (red) respectively. The dashed lines show the best fit to the data.

beyond what was stated previously and use the possibility to vary the confinement size in our simulations to find a quantification for the crossover polymer length.

Due to the reflecting boundaries, polymer segments far apart along the polymer become uncorrelated in their motion when the polymer in between touches the boundary. Therefore the confinement leads polymer subchains to become independent when the polymer is long enough compared to the simulated cell to frequently encounter the boundary. To quantify this, we compute how many polymer beads lie on average between two encounters with the boundary. For a given box size we construct random configurations of very long polymers. Polymers are chosen long enough so that the measured quantities are independent of polymer length. There are two different quantities that can be computed. First, the average length of polymer subchains can be computed, if every encounter with the boundary would cut the chain. On the other hand, we can select a few binding sites along the polymer and compute the average length of the subchains with a binding site, which we call N_{boundary}. This means that starting from the selected binding site, we move to the left and to the right along the polymer and count the number of beads, until a monomer lies at the boundary. Interestingly, the two quantities differ considerably (see fig. D.3b), which can be explained by the waiting time paradox [171], whenever the distribution of subchain lengths does not follow a Poisson distribution. When selecting a subchain by choosing the position of the

binding site, a subchain is not chosen randomly but with a probability proportional to its size and therefore the average size of the subchain with the binding site is larger than the mean subchain size. The average subchain length of the binding site is shown in fig. III.3a by the blue data. We observe quite a good agreement of the crossover chain length with N_{boundary}. For polymers larger than N_{boundary}, the average correlated length is still N_{boundary}, so that additional beads do not influence the search time, thus leading to the observed plateau.

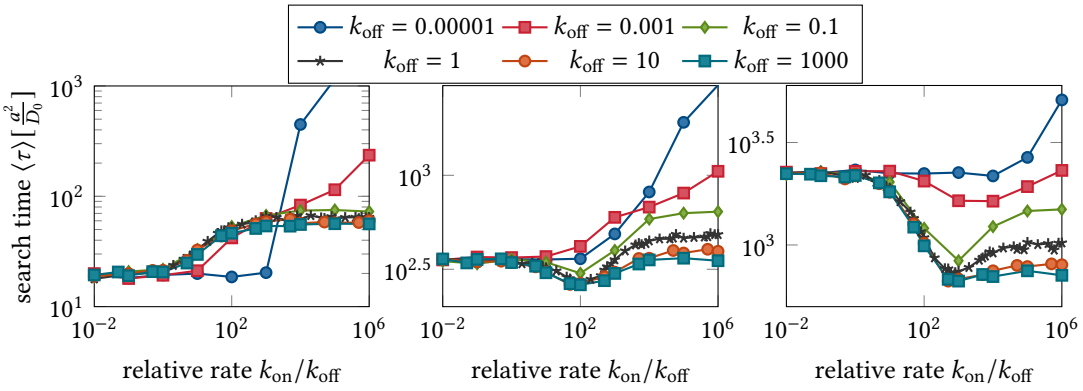
2.1.4 Saturation search time

For very long polymers compared to the confinement it can be assumed that the binding site movement is dominated by subdiffusive motion, because there is no net displacement of the polymer centre of mass. The scaling of the mean first passage time of a subdiffusively moving random walker to reach a target inside a domain V with reflecting boundaries has been computed in a previous study [170]. In particular, the MFPT averaged over the initial position has been obtained in the large volume limit to scale as $\langle \tau \rangle \propto V^{d_w/d_f}$ in the case of compact exploration $d_w > d_f$. Here, d_w is the walk dimension, defined by the mean squared displacement $\langle \boldsymbol{r}^2(t) \rangle \propto t^{2/d_w}$ and d_f is the fractal dimension, such that the linear size of the medium scales as $L \propto V^{1/d_f}$, i. e. $d_f = 3$ in our case. For the subdiffusive motion of the binding site it follows that $d_w = 4$ and we expect the saturation search time to scale as $\langle \tau_{\text{plateau}} \rangle \propto L^4$. For non compact exploration $d_w < d_f$ it was found that $\langle \tau \rangle \propto V$, which is the well-known result for normal diffusion ($d_w = 2$, $d_f = 3$). Since for non compact exploration the number of visited sites grows linearly with the number of steps, the MFPT is proportional to the volume. For compact exploration however each site is on average oversampled and the mean number of sites visited grows slower than linearly with the number of steps, leading the MFPT to increase faster than linearly with the volume.

As demonstrated in fig. III.3b in red, the scaling $\langle \tau_{\text{plateau}} \rangle \propto L^4$ is obtained when the binding site is placed at one end of the polymer. However, for the binding site at the middle of the polymer, the saturation search time increases even steeper with box size, as shown in blue. Deviations could arise because the scaling argument assumes a Markovian random walker, whereas a monomer inside a polymer has non-Markovian properties [172], because it depends on the interactions of the other monomers in the chain. Moreover, the protein is assumed to lie within the confinement V, whilst it is placed at the boundary in our simulations. This could have a larger effect when the binding site is a mid monomer, because it is kept away from the boundary by the two polymer strands it is connected to, while an end monomer is less restricted in its movement.

2.2 Non-Specific Binding

So far, the target search of a single binding site at the middle of the polymer for a single static protein at the surface has been discussed. Now we allow the protein to move with a constant rate k_{2d} and vary the rates k_{on} and k_{off} for non-specific binding and unbinding of the protein to the polymer independently to analyse the effect on the mean search time.



(a) High polymer density **(b)** Medium polymer density **(c)** Low polymer density

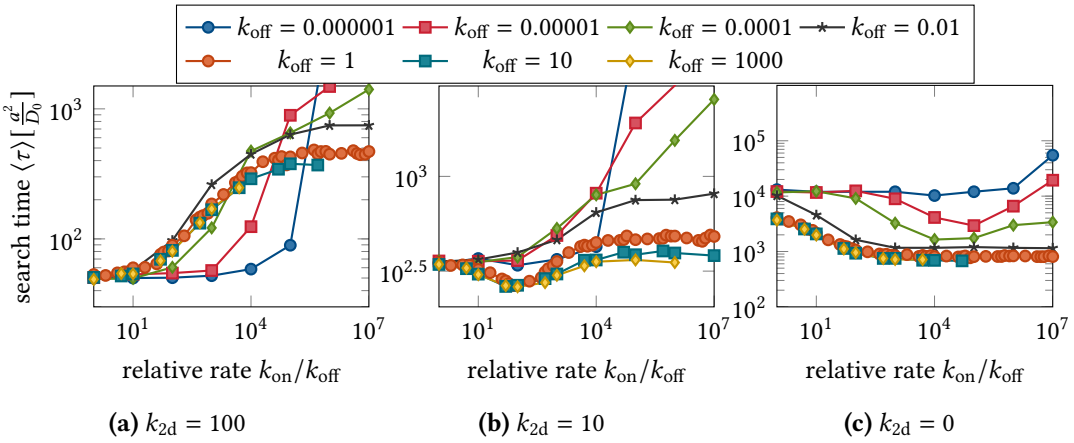
Figure III.4: Non-specific binding and unbinding. Mean search time $\langle\tau\rangle$ as a function of the relative binding rate k_{on}/k_{off} for different values of k_{off} without sliding. (**a**) For high polymer density ($L = 5$, $N = 100$) the search time increases for increasing binding rate and reaches a higher plateau. (**b**) For intermediate polymer density ($L = 10$, $N = 100$) the search time forms a minimum for large unbinding rates. The search time minimum is at $k_{on}/k_{off} \approx 100$. (**c**) For low polymer density ($L = 15$, $N = 100$) the search time minimum is shifted to $k_{on}/k_{off} \approx 1000$ and for very fast unbinding a lower search time plateau is formed.

2.2.1 Optimum of the relative binding rate

In fig. III.4 the mean search time is shown as a function of the relative binding rate k_{on}/k_{off} for varying unbinding rates. Sliding is not included in the shown simulations but we will comment on its effect later. The behaviour depends on the box size L, the polymer length N and the protein diffusion rate k_{2d}, the latter being discussed below. Hence the mean search time is shown for constant protein diffusion and different polymer densities, i.e. increasing box size or decreasing polymer length respectively.

Figure III.4a shows the mean search time as a function of the relative binding rate for very high polymer density (very small boxes or long polymers). Due to the crowded environment for high polymer density it is very likely that the binding site is on a polymer segment that is uncorrelated from the bound segment. Therefore transient binding does not speed up the search process. Instead, it only hinders 2d diffusion of the protein and therefore slows down the search. The search time is unaffected for very small binding rates and then increases monotonically with k_{on}/k_{off} and reaches a higher plateau. While the initial behaviour is independent of the unbinding rate, because the binding rate is so small that it does not have an effect and therefore also no unbinding occurs, it affects the following course. Especially for very small unbinding rates the search time increase is very steep and the second plateau is high. Binding is approximately permanent in this regime of very small unbinding rates and therefore it dramatically slows down the search process when sliding is disabled, as it hinders polymer motion without bringing the specific binding site closer to the protein.

For medium polymer density, depicted in fig. III.4b, the transition towards a higher plateau exists only for approximately permanent binding (small unbinding rates). For fast unbinding



(a) k_{2d} = 100 **(b)** k_{2d} = 10 **(c)** k_{2d} = 0

Figure III.5: Protein diffusion rate dependence of non-specific binding and unbinding. Mean search time $\langle\tau\rangle$ as a function of the relative binding rate k_{on}/k_{off} for different values of the protein diffusion rate k_{2d}. Sliding was not included in the shown simulations. (**a**) For fast protein diffusion (k_{2d} = 100) the search time increases for increasing binding rate and reaches a higher plateau. (**b**) For intermediate protein diffusion rate (k_{2d} = 10) the search time forms a minimum for large unbinding rates at $k_{on}/k_{off} \approx 100$. For smaller unbinding rates the mean search time increases monotonously with the relative binding rate and transitions from a lower to a higher plateau. (**c**) For an immobile protein (k_{2d} = 0) the search time minimum is shifted to $k_{on}/k_{off} \approx 100000$. For very fast unbinding rates the mean search time decreases monotonously with the relative binding rate and transitions from a higher plateau to a lower plateau.

a shallow minimum is formed at $k_{on}/k_{off} \approx 100$, before the curve increases again to reach the higher plateau. In this regime of larger unbinding rates, where binding is not permanent, even without sliding of the protein along the polymer it is favourable for the search process that the protein binds non-specifically to the polymer, as demonstrated by the search time minimum. Transient binding tethers the polymer to the surface for some time so that the binding site, when it is close by, might reach the protein faster. However, there is an optimal time the protein spends being bound to reach the smallest target search times.

As shown in fig. III.4c, for very low polymer density a third regime appears. For very fast unbinding rates no minimum is formed, but a lower search time plateau. Since the polymer is small compared to the box size, transient binding always brings the binding site close to the protein. Therefore the search time minimum is shifted towards larger values of $k_{on}/k_{off} \approx 1000$ and especially very short binding periods speed up the process. The higher search time plateau now appears only for very small unbinding rates, i.e. approximately permanent binding.

Enabling sliding of the protein along the polymer does not affect the location of the search time minimum (see fig. D.4). However it has a major effect on the two plateaus: instead of increasing with the binding rate, the search time always decreases and reaches a lower plateau, showing that binding in general is favourable when sliding is enabled.

Decreasing the protein diffusion rate k_{2d} yields a very similar behaviour as for decreasing polymer density, shown in fig. III.5. When the protein is diffusing very fast on the surface, the

efficient search of the protein is only slowed down by non-specific binding and increasing the relative binding rate leads the mean search time to reach a higher plateau. For intermediate protein diffusion, transient binding leads to the formation of a minimum. While transient tethering of the protein to the polymer speeds up the procedure in this domain, the duration of the binding events has to be limited to a certain value. For very slow protein diffusion or static proteins, binding that is not permanent leads to a lower plateau for larger relative binding rate, because the protein cannot be slowed down by tethering and therefore the chance of bringing the binding site closer to the protein speeds up the process.

2.3 Protein Diffusion

We consider the effect of protein diffusion on the surface of the enclosing volume and along the polymer. To this end, the mean search time is computed for varying diffusion rate k_{2d} and sliding rate k_{1d} respectively.

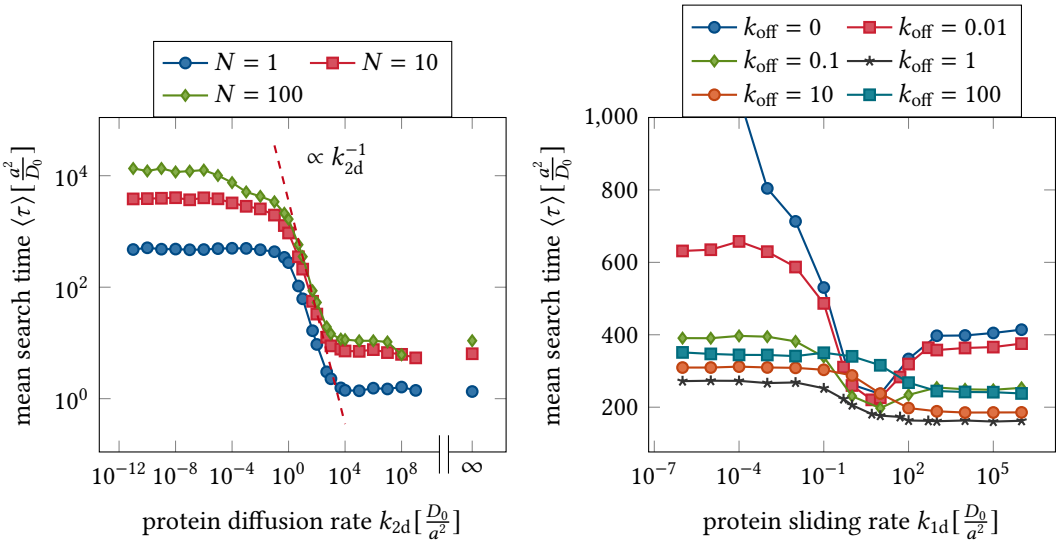
2.3.1 For intermediate protein diffusion rates the search process is limited by protein movement

To focus on the effect of 2d diffusion along the surface, we simulated a target search without non-specific binding and sliding. In fig. III.6a the mean search time is shown for varying protein diffusion on the surface and different polymer lengths. When the protein is diffusing very slowly, it has almost no impact on the search time and we observe a plateau. The larger the polymer, the earlier starts a visible effect on the search time, because once the binding site is relatively close to the protein, it spends more time in the compact search mode if it is part of a longer polymer and therefore even a slowly diffusing protein appears larger than an immobile protein. At around $k_{2d} = 1$, independent of polymer length, the mean search time starts to decrease inversely proportional to k_{2d}, showing that the search process is strongly limited by the protein diffusion. At around $k_{2d} = 10^3$ a lower plateau is reached. In this regime the protein is so fast that it covers the whole surface, leading to the same result as if the whole surface was the target. This is shown by the agreement with the mean search times for $k_{2d} \to \infty$ in fig. III.6a, which were obtained by placing a static protein at each lattice point on the surface.

2.3.2 Sliding rate optimum

The mean search time as a function of the sliding rate is depicted in fig. III.6b for different values of the unbinding rate and constant binding rate. The behaviour can be separated into two regimes, depending on the unbinding rate. For very small unbinding rates $k_{off} \leq 0.1$ unbinding events are rather unlikely and we call this domain the “permanent binding” regime. In contrast, for larger unbinding rates $k_{off} \geq 1$, binding is only temporary and we refer to this regime as “transient binding”.

For transient binding ($k_{off} \geq 1$) we observe a similar behaviour as for the 2d diffusion rate in fig. III.6a. After an initial regime, where the sliding rate is too slow to affect the search time, an intermediate domain is reached where the mean search time decreases with the sliding rate, before a lower plateau is formed. When the intermediate area is reached depends on the



(a) Effect of protein diffusion along the surface

(b) Effect of sliding along the polymer

Figure III.6: Protein diffusion. (**a**) Mean search time $\langle\tau\rangle$ as a function of the protein diffusion rate k_{2d} for different polymer sizes N. Between a higher plateau for slow protein diffusion and a lower plateau for fast protein diffusion the search time varies inversely proportional with k_{2d}. The data points for $k_{2d} \to \infty$ were obtained by placing a protein at each lattice site on the surface. (**b**) Mean search time $\langle\tau\rangle$ as a function of the protein sliding rate k_{1d} for different values of the unbinding rate and constant $k_{on} = 100$. While for large unbinding rates the search time varies only slightly between a higher plateau for slow sliding and a lower plateau for fast sliding, for small unbinding rates the search time reaches a minimum at around $k_{1d} \approx 1$.

unbinding rate. The smaller it is the earlier starts the search time decrease because the protein spends more time being bound and therefore the sliding rate has a larger effect. Since the protein can only slide along polymer segments that are at the surface, the average distance that it travels while sliding is not very large. Between the two search time plateaus there is only a factor of 1.66, meaning that an infinitely fast sliding protein appears less than twice as big as a protein that is not sliding. Assuming one polymer bead is located at a flat surface, the probability that the neighbouring bead is at the surface is $p = 8/17$ for a random walk with diagonal bonds. This leads to a mean number of segments at the boundary of $\langle n\rangle \approx 1.68$, in good agreement with the factor between the two plateaus. This factor is highly conserved for varying polymer length and protein diffusion rate (see figs. D.5 and D.7). It only increases for very small boxes (2.33 for $L = 5$, see fig. D.6) because at corners and edges the probability for a neighbouring bead to stay at the surface is larger than assuming a flat surface, as was done for the estimate above.

The almost constant sliding range, independent of most parameters differs from the sliding range ("antenna length") in the facilitated diffusion model for cytoplasmic proteins, which was found to scale with the square root of the diffusion constant of sliding D_{1d} and the average time of a sliding event τ_{1d}, yielding $\bar{n} = 2\sqrt{D_{1d}\tau_{1d}}$ [37]. However, experimentally obtained

diffusion coefficients suggest, that a protein can slide over a range of 100–1000bp [37]. The length of 1.66 Kuhn segments of DNA ($l_K \approx 100\,\text{nm} = 300\,\text{bp}$) corresponds to 498 bp, which agrees well with the experimentally calculated sliding range.

For approximately permanent binding ($k_{off} \leq 0.1$) the behaviour changes and we observe an optimal sliding rate, where the mean search time reaches a minimum. As before, very small sliding rates lead to a higher plateau, then a minimum is formed before the lower plateau is reached. The minimum appears only for even smaller unbinding rates if the binding rate is smaller (see fig. D.8a) because the relative rate k_{on}/k_{off} determines how much time the protein spends being bound and there is only an optimum for approximately permanent binding.

This behaviour is similar to what is observed for cytoplasmic proteins reaching a binding site on a polymer: since the 1d search is more redundant than 3d diffusion, there is an optimum how much time the protein spends in each mode. In an optimal search the protein spends half of its time sliding along the polymer and half of its time diffusing in free space [44].

However, it is not clear what leads to the search time minimum in our case. Since the minimum is also observed for $k_{2d} = 0$ it is most likely not a competition between 2d diffusion and sliding. Moreover the minimum is only present for approximately permanent binding, i.e. 2d diffusion cannot play a major role. The position of the minimum at around $k_{1d} \approx 1$ is independent of all parameters and therefore solely depends on the polymer diffusion rate. When a permanently bound protein slides very fast, it might tether a large fraction of the polymer to the surface and prevent free diffusion of the polymer. While this also happens for transient binding, for permanent binding more or less the same portion of polymer stays tethered to the surface. Also dragging of the protein is reduced when the polymer moves much slower than the protein moves by sliding. However, this is also the case for shorter binding periods, where very fast sliding does not lead to a search time increase.

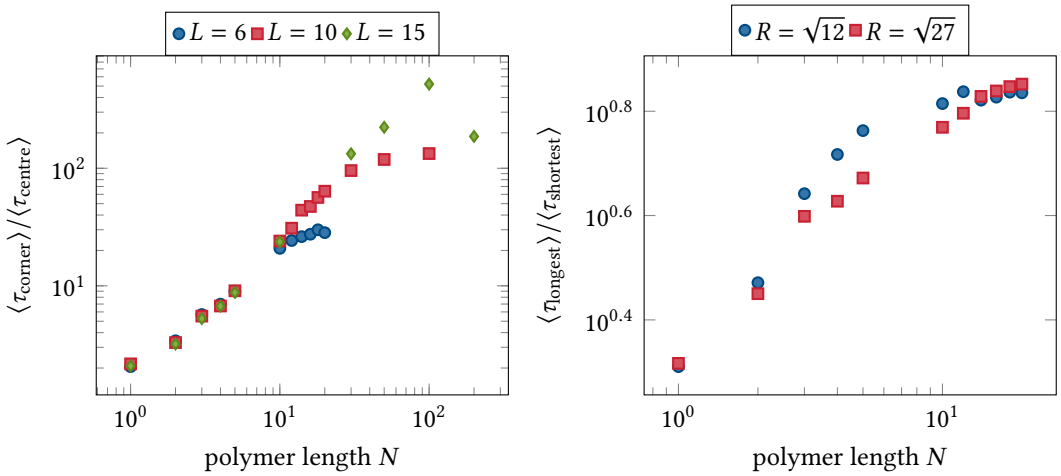
There is no search time minimum for small boxes or small polymers, i.e. when the binding site is close in space to the bound polymer bead. In this domain it is more likely for the target search to be successful even when polymer motion is impeded. For short polymers the binding site is close to the bound site along the polymer contour, making it more likely for the protein to locate the binding site by sliding. For small boxes the binding site is close to the bound site in space and more likely to be at the surface, also increasing the chance of being found by sliding.

2.4 Geometry of the Search Volume

All simulations shown so far have been performed inside a cubic box. We now consider the effect that this geometry has on the target search and compare to simulations inside an approximately spherical volume.

2.4.1 Polymer length dependent corner effect

When a polymer is confined such that it is smaller than its natural mean size, i.e. its radius of gyration R_g, the reduction of the number of accessible conformations leads to a reduction of the conformational entropy and an increase in the free energy of the polymer. We therefore expect a certain exclusion effect of the polymer from the corners of the cubic box. To quantify



(a) Cubic volume **(b)** "Spherical" volume

Figure III.7: Corner effect. (**a**) Ratio of the mean search time to reach a static protein at the corner $\langle \tau_{\text{corner}} \rangle$ and the mean search time to reach a static protein at the face centre $\langle \tau_{\text{centre}} \rangle$ of the cubic volume as a function of polymer length and for different box sizes. (**b**) Ratio of the longest and the shortest mean search time to reach a static protein as a function of polymer length and for different sizes of the spherical volume.

this, we plot the ratio of the mean search time to reach a static protein at the corner $\langle \tau_{\text{corner}} \rangle$ and the mean search time to reach a static protein at the face centre $\langle \tau_{\text{centre}} \rangle$ as a function of polymer length and for different sizes of the cubic box in fig. III.7a. The binding site was placed at the middle of the polymer. While for monomers the corner effect is very small, it increases dramatically with polymer length and independent of the box size. The polymer length dependence arises because the longer the polymer is, the stronger it is confined by the corner. Only for long polymers compared to the box size the ratio stops increasing and reaches a plateau, which is the same effect that was discussed before. Figure III.7b shows the corresponding plot for an approximately spherical confinement of different volume. Since it is not obvious which of the possible locations of the static protein is most easily reachable and the least reachable respectively, corresponding to the the face centre and the corner of the cube, we plot the ratio between the longest and the shortest mean search time. While different locations of the static protein still lead to different mean search times even for an approximately spherical volume, the effect is considerably smaller than for the cube. The ratio of search times is still increasing with polymer length, but much less steep compared to the cubic volume.

2.4.2 Quantification of the corner effect

Since the corner effect arises due to a reduction of the conformational entropy and therefore an increase in the free energy of the polymer compared to a free polymer, we attempt to explain its extent by approximating the free energy of localising a certain monomer at a certain position inside the cubic box. The probability that the binding site lies in a subvolume

δV is given by

$$\mathcal{P}(\delta V) = \left(\frac{\delta V}{V}\right) \exp[-\beta \Delta F_{\mathrm{loc}}] \,,$$

with the localisation free energy ΔF_{loc} and the inverse temperature $\beta = 1/k_{\mathrm{B}}T$. If we consider the free energy of localising an end monomer at the corner of a cubic box, it can be approximated by the free energy difference between a free polymer (0) and a polymer with the end tethered to a corner (c) of an infinitely large box [86]

$$\begin{aligned} \Delta F_{\mathrm{loc}} &\approx \Delta F_{0\mathrm{c}} = F_0 - F_{\mathrm{c}} \\ &= \beta^{-1} \ln \mathcal{Z}_{\mathrm{c}} - \beta^{-1} \ln \mathcal{Z}_0 \,, \end{aligned}$$

where the equality holds in the limit $V \to \infty$. It is well known, that the partition function $\mathcal{Z}$ of a random walk and a self-avoiding walk can be written as

$$\mathcal{Z} = z^N N^{\gamma - 1} \,,$$

where z depends on the particular lattice and the critical exponent γ depends only on the dimensionality and the topology of the polymer. With this we arrive at

$$\Delta F_{\mathrm{loc}} \approx \beta^{-1} \alpha_{0\mathrm{c}} \ln N$$

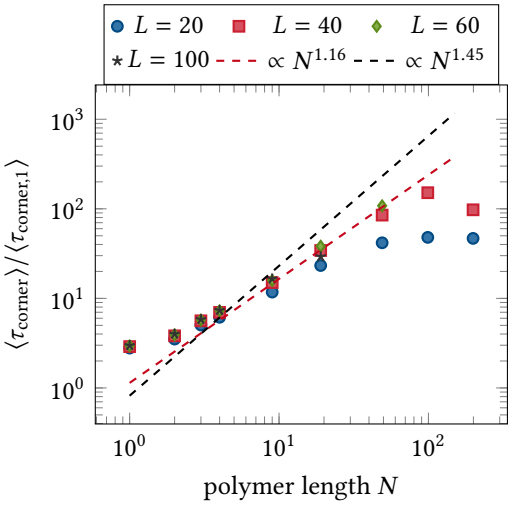
with $\alpha_{0\mathrm{c}} = \gamma_{\mathrm{c}} - \gamma_0$. If we assume that the mean first passage time of a certain monomer to reach a certain point in the box is inversely proportional to the probability of this configuration, we obtain

$$\langle \tau \rangle \propto \mathcal{P}(\delta V)^{-1} \propto N^{\alpha} \,.$$

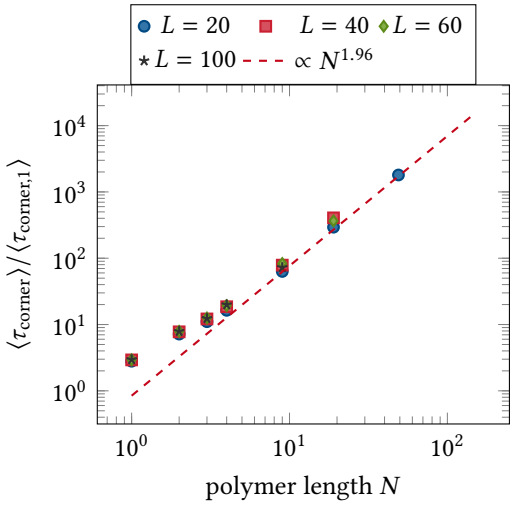
Approximate values for the scaling exponents α have been obtained by Monte Carlo simulations estimating the change in the internal conformational free energy upon localisation of an end monomer to either an infinite hard wall or the corner of an infinitely large cube [86]. The exponents obtained for a 3d self-avoiding walk are summarised in table III.1. Here, "c" stands for localising a monomer at the corner, "f" stands for localising a monomer at the face centre, "s" stands for localising a monomer at a sphere and "mid" refers to the mid monomer being localised instead of the end monomer.

In the same study Monte Carlo simulations of an off-lattice hard sphere bead-spring model were used to estimate the exponents for finite confinement sizes. While smaller volumes in general led to smaller exponents, they found already a good agreement with the infinite volume exponents for $L/R_{\mathrm{g}} \approx 15$. Those simulations were also used to estimate the exponent for the localisation of the mid monomer at the boundary of a sphere, which is shown for $L/R_{\mathrm{g}} \approx 10$ in table III.1. The localisation of an end monomer at a sphere for $L/R_{\mathrm{g}} \approx 10$ is given by $\alpha_{0\mathrm{s}} \approx 0.54$, close enough to $\alpha_{0\mathrm{f}}$ so that we use $\alpha_{0\mathrm{s}}^{\mathrm{mid}}$ as an estimate for the localisation of the mid monomer at the face centre of the cube.

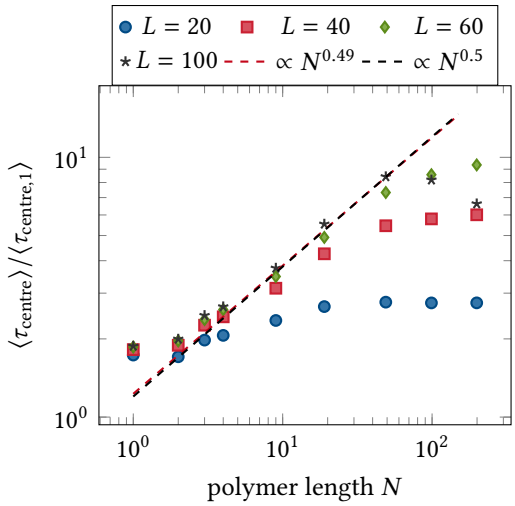
While the critical exponents γ may differ between 3d random walks and self-avoiding walks, it was found [173] that their difference α is in general very similar. We therefore use the exponents obtained for self-avoiding walks to compare to our simulation results.



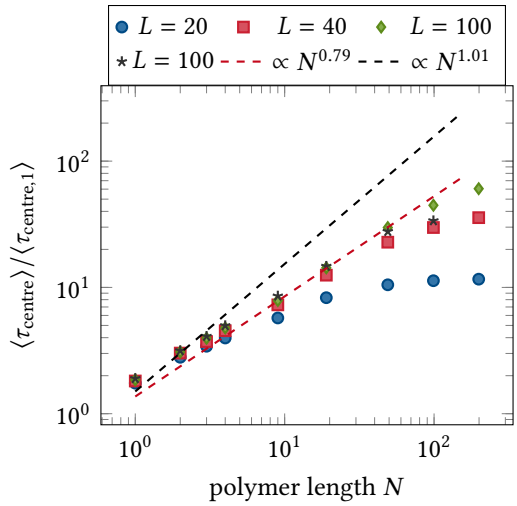
(a) End monomer reaching the corner



(b) Mid monomer reaching the corner



(c) End monomer reaching the face centre



(d) Mid monomer reaching the face centre

Figure III.8: Scaling exponents. The red dashed lines show the best fit for the largest simulated box size, while the black dashed lines show the expected critical exponents according to table III.1. (**a**) Mean search time of the end of the polymer to reach the corner of the box as a function of polymer length for different box sizes. (**b**) Mean search time of the middle of the polymer to reach the corner of the box as a function of polymer length for different box sizes. (**c**) Mean search time of the end of the polymer to reach the face centre of the box as a function of polymer length for different box sizes. (**d**) Mean search time of the middle of the polymer to reach the face centre of the box as a function of polymer length for different box sizes.

In fig. III.8 the mean search time of an end monomer or a mid monomer respectively to reach the corner or the face centre of the cubic box is shown. The search time is normalised by the mean search time of a free monomer and plotted as a function of polymer length and for different sizes of the cubic volume. The red dashed lines show the best fit to $\langle\tau\rangle(N) = N^c$ with the fit parameter c (see table D.2), the black dashed lines show $\langle\tau\rangle(N) = N^\alpha$ with the exponents from table III.1. $\langle\tau\rangle \propto N^x$ indeed holds for the regime before the search time reaches a plateau. For the end monomer reaching the face centre in fig. III.8c there is quite a good agreement with the expected exponent for the largest simulated volume $L = 100$. For the end monomer reaching the corner in fig. III.8a we find a smaller exponent than expected, which might be due to the fact that the simulations for $L = 100$ were too slow to finish. The largest deviations are observed for the mid monomer reaching the face centre of the cube in fig. III.8d. Since to our knowledge there is no previous data on this exponent, we compared it to the exponent for the mid monomer reaching a sphere. Although for the end monomer the exponents for reaching the face centre of a cube and for reaching a sphere are very similar, it might be that this is not the case for the mid monomer, therefore leading to the large deviations. In a previous study [86] a good agreement with the infinite volume estimates has been found for $L/R_g \approx 15$. For $L = 60$ and $N = 100$ we obtain $L/R_g \approx 21$, which is why we do not expect finite size effects to play a major role. However, in the previous study off-lattice simulations have been used, which could lead to an earlier convergence of the exponents.

Random walk:	α_{0f}		
	0.5		
Self-avoiding walk:	α_{0f}	α_{0c}	α_{0s}^{mid}
	0.47	1.45	1.01

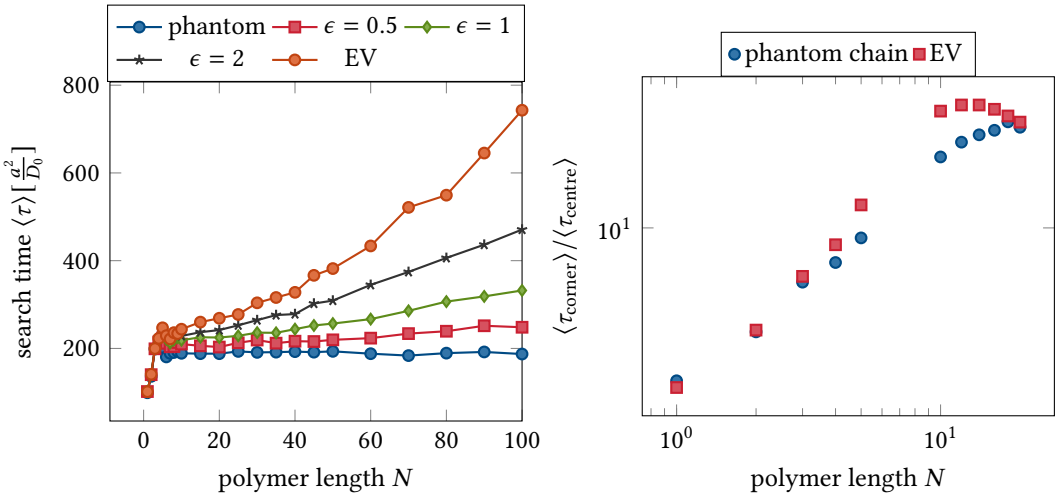
Table III.1: Scaling exponents. Exponents for the 3d random walk and self-avoiding walk are taken from [86] and references therein.

2.5 Excluded Volume Interactions

So far, all shown simulations have used a phantom chain implementation of the polymer. We now investigate the effect of excluded volume interactions of the polymer on the target search process. Moreover, partial self-exclusion of polymer beads and bonds is considered to transition between a phantom chain and a self-avoiding chain.

2.5.1 Search time diverges after reaching the plateau

The mean search time as a function of polymer length is shown in fig. III.9a for different values of the energy penalty ϵ, from $\epsilon = 0$ (phantom chain) to $\epsilon = \infty$ (self-avoiding chain). Excluded volume interactions of the polymer prevent both beads and bonds to occupy any lattice site more than once. For partial excluded volume this constraint is softened but penalised with an energy penalty ϵ. With excluded volume interactions between polymer beads and bonds the mean search time first increases linearly, and then reaches a plateau for a small range of N before it diverges. For partial self-exclusion we observe the same effect but the increase for



(a) Search time plateau

(b) Corner effect

Figure III.9: Excluded volume effect. (**a**) Mean target search time $\langle\tau\rangle$ as a function of polymer length N for different values of the energy penalty ϵ, ranging from $\epsilon = 0$ (phantom chain) to $\epsilon = \infty$ (self-avoiding chain). The static protein is placed at the face centre of the cubic volume, the polymer is initialised with a random walk and the mean search time is taken over $\geq 10^3$ runs. With excluded volume interactions, the search time increases further and diverges after reaching the plateau, the increase being faster the larger the energy penalty is. (**b**) Fraction between the search time with the static protein at the corner and the search time with the protein at the face centre of the cubic box for a phantom chain (blue) and a self-avoiding chain (red). The corner effect is slightly stronger with self-exclusion than for a phantom chain.

long polymers becomes less steep the smaller the energy penalty. It therefore transitions into the curve for phantom chains where the increase for long polymers is replaced by a plateau. The same plot for larger volumes is shown in fig. D.8b.

Figure III.9b shows the ratio of the mean search time to reach a static protein at the corner of the cubic box and the mean search time to reach a static protein at the face centre as a function of polymer length for a phantom chain (blue) and a self-avoiding chain (red). The ratio increases a bit faster with self-exclusion, showing that the corner effect is stronger than for a phantom chain. This behaviour is expected because the mean size of the swollen self-avoiding chain is larger than the size of a phantom chain with the same chain length and therefore it experiences a stronger exclusion from the corners.

2.6 Number of Proteins and Binding Sites

While so far we have only covered target search processes with a single protein on the surface and a single binding site on the polymer, we shortly comment on the effect of increasing the number of either of the two.

2.6.1 Linear dependence on the number of proteins and binding sites

The mean search time is inversely proportional to the number of binding sites when the number of proteins is kept constant and the binding sites are placed far from each other along a circular polymer, shown in fig. III.10a. This is the expected behaviour when the different binding sites can be considered as independent searchers. When the number of binding sites becomes so large that more than one is located on one of the independent polymer segments, the search time decreases less steep than inversely proportional with the number of searchers. The same can be observed when the binding sites are placed next to each other along the polymer.

When multiple static proteins are placed randomly on the surface of the confinement, with a single binding site on the polymer, the mean search time first decreases steeper than inversely proportional to the number of proteins, as shown in fig. III.10b in blue. This is an effect that arises because the corners are more difficult to reach for the polymer than for example the face centre. The more proteins are placed randomly, the more likely there is one that is not at a corner and thus easily reachable. Therefore the mean search time first decreases steeper and then inversely proportional with the number of proteins. The decrease slows down dramatically when a large fraction of the surface is covered with proteins. This is underpinned by placing the static proteins next to each other at the face centre of the cube, corresponding to the red data in fig. III.10b. Now the dependence on the number of proteins is initially linear but becomes less steep as the proteins approach the corners of the cube.

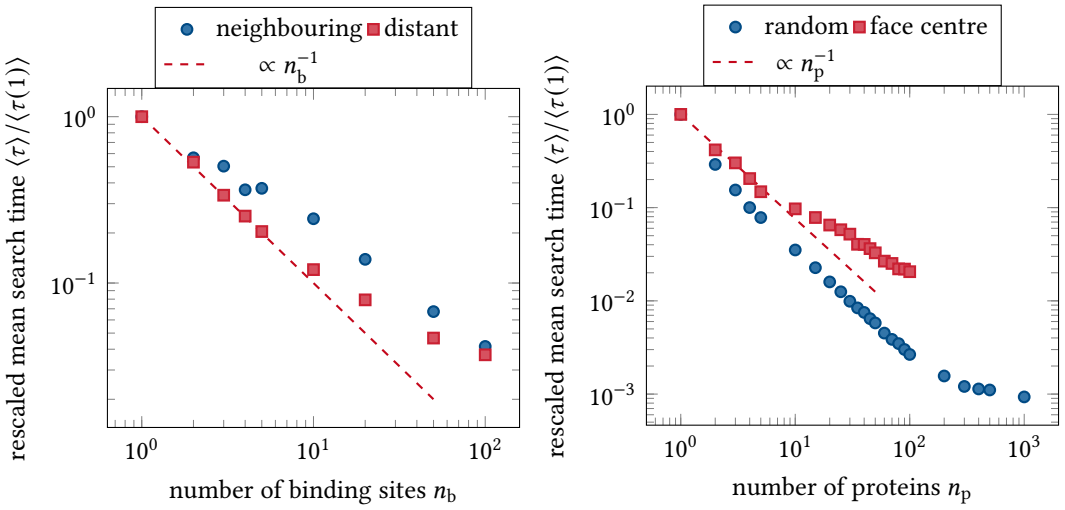
2.7 Search Time Distribution

After extensively discussing the target search process in terms of the mean search times, we will now consider the search time distribution and what determines its shape. To this end the probability density function of the search time $\Psi(\tau)$ is computed from $\geq 10^3$ simulation runs with varying protein diffusion rate without non-specific binding and sliding. Plots of the cumulative distribution function (CDF) $\Phi(\tau) = \int_0^\tau \Psi(t')\mathrm{d}t'$ for a randomly placed protein are depicted in fig. III.11 together with the best fit to a single exponential distribution, shown in red. Depending on the protein diffusion rate different distributions are observed. As we will see in the following section, the distribution also depends on whether the protein is placed at a fixed or randomly chosen position on the surface.

2.7.1 Static protein leads to single exponential distribution

For a static protein, placed at a fixed position e.g. at the face centre of the box, the distribution agrees well with a single exponential (see fig. D.10a), which is the well-known behaviour for a Brownian particle reaching a small target [174] and has also been reported for polymers searching for a static target in confinement [85]. Since the protein is immobile, the search process depends solely on the diffusing polymer and can be described by a single transition rate k_1^+. It therefore can be explained by a two state process

$$S_1 \xrightarrow{k_1^+} S_2\,,$$



(a) Multiple binding sites **(b)** Multiple proteins

Figure III.10: Multiple binding sites and proteins. (**a**) Mean search time $\langle\tau\rangle$ rescaled with the search time for a single binding site $\langle\tau(1)\rangle$ as a function of the number of binding sites. The binding sites are either placed next to each other (blue) or as far apart from each other as possible (red) along the circular polymer. (**b**) Mean search time $\langle\tau\rangle$ rescaled with the search time for a single static protein $\langle\tau(1)\rangle$ as a function of the number of proteins. The proteins are placed either randomly on the surface (blue) or next to each other on the face centre of the cube (red).

where S_1 corresponds to the initial, unbound state and S_2 is the state where the binding site is bound to the static protein. The probability density function $\Psi(\tau)$ and the cumulative distribution function $\Phi(\tau)$ are given by (see appendix C1)

$$\Psi(\tau) = k_1^+ e^{-\tau k_1^+},$$
$$\Phi(\tau) = 1 - e^{-\tau k_1^+}.$$

with the mean first passage time

$$\langle\tau\rangle = \frac{1}{k_1^+}.$$

However, when the protein is immobile or very slowly diffusing (fig. III.11a and fig. III.11b), and randomly placed on the box surface, the geometry of the cubic simulation box leads to a deviation from the single exponential distribution. As we have seen above, the process is well described by a single exponential search time distribution with a certain rate when the static protein is placed at a fixed position. However, since this rate depends on the specific position of the protein on the surface, placing the protein randomly and averaging over several simulations leads to an overlay of multiple single exponential processes with different rates. It therefore requires multiple exponentials to describe the process, which we call a simultaneous multi step process.

In the case of two different positions of the static protein, it can be described by a three state process with mixed initial condition, where a fraction x starts in state S_3 and a fraction $1-x$ starts in state S_1,

$$S_1 \xrightarrow{k_1^+} S_2 \xleftarrow{k_3^-} S_3 \,.$$

S_1 and S_3 correspond to the unbound state with the two possible positions of the protein and S_2 is the final bound state. This leads to a probability distribution (see appendix C2)

$$\Psi(\tau) = x k_3^- e^{-k_3^- \tau} + (x-1) k_1^+ e^{-k_1^+ \tau} \,,$$
$$\Phi(\tau) = 1 - x e^{-k_3^- \tau} + (x-1) e^{-k_1^+ \tau} \,,$$

with a mean first passage time

$$\langle \tau \rangle = \frac{x}{k_3^-} + \frac{1-x}{k_1^+} \,.$$

While a weak form of this effect is also visible for approximately spherical cells, for the cubic box it arises mainly due to exclusion of the polymer from the corners.

That the deviation from the single exponential distribution indeed arises due to the corner effect can be shown by simulating a polymer of length $N = 1$, where the corner effect is so weak that for slowly diffusing proteins the distribution is always a single exponential (see fig. D.9).

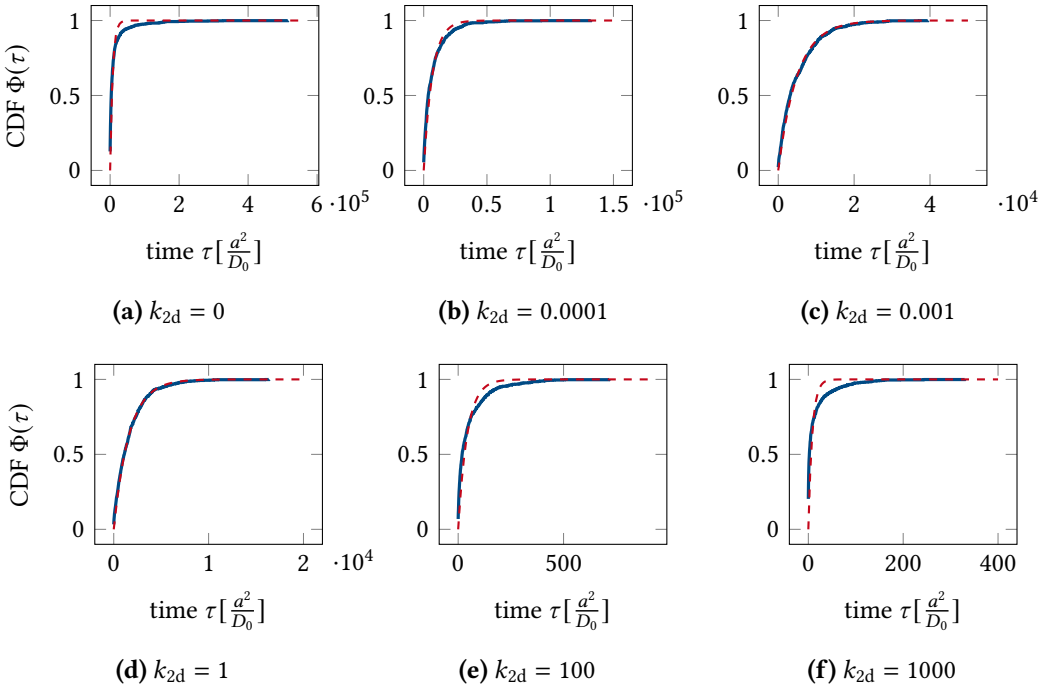
When the protein is diffusing still slowly, such that the polymer movement dominates, but fast enough to escape the corners (fig. III.11c and fig. III.11d), the corner effect is reduced and the distribution can be well fitted by a single exponential.

2.7.2 Moving protein leads to sequential two step process

For faster diffusing, randomly initialised proteins (fig. III.11e and fig. III.11f), a different effect leads again to a deviation from the single exponential distribution. That the corner effect is not responsible for the deviation can again be shown by simulating a polymer of length $N = 1$, where the corner effect is so weak that for slowly diffusing proteins the distribution is always a single exponential but for faster protein diffusion the deviation from the single exponential still persists (see fig. D.9). Since both polymer movement and protein diffusion significantly contribute to the search process, there are two relevant rates in the system and we observe the distribution of a reversible sequential two step process

$$S_1 \underset{k_2^-}{\overset{k_1^+}{\rightleftharpoons}} S_2 \xrightarrow{k_2^+} S_3 \,.$$

State S_1 corresponds to the initial unbound state and state S_3 to the final state where the binding site is bound to the protein. The intermediate state S_2 could correspond to configurations where the binding site is already close to the surface, but not yet bound to the protein.



(a) $k_{2d} = 0$ **(b)** $k_{2d} = 0.0001$ **(c)** $k_{2d} = 0.001$

(d) $k_{2d} = 1$ **(e)** $k_{2d} = 100$ **(f)** $k_{2d} = 1000$

Figure III.11: Cumulative distribution. (**a**)-(**f**) Cumulative distribution function $\Phi(\tau)$ of the first passage time for varying diffusion rate of the protein. The red dashed lines show the best fit to an exponential distribution. Three regimes can be observed: while for a very slow and very fast protein the CDF deviates from an exponential distribution, they agree very well for $k_{2d} = 0.01$–1.

Assuming the system to be initially in state S_1, this leads to the first passage time distribution (see appendix C1)

$$\Psi(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}}{\alpha - \beta}$$
$$\Phi(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}}}{\alpha - \beta},$$

with

$$\alpha = 2\left(k_1^+ + k_2^+ + k_2^- - \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+ k_2^+}\right)^{-1},$$
$$\beta = 2\left(k_1^+ + k_2^+ + k_2^- + \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+ k_2^+}\right)^{-1},$$

and the mean first passage time is given by

$$\langle\tau\rangle = \alpha + \beta = \left(1 + \frac{k_2^-}{k_2^+}\right)\frac{1}{k_1^+} + \frac{1}{k_2^+}.$$

The above distribution assumes that the system is initially in state S_1. However, for random initialisation it could also be that for some realisations of the search process the system starts already from state S_2, e.g. with the binding site being already at the surface of the confining volume. When a mixed initial condition is assumed, with a fraction x starting from state S_2 and a fraction $1-x$ starting from state S_1, the probability distribution has the form

$$\Psi(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}} + c\left(\alpha e^{-\frac{\tau}{\beta}} - \beta e^{-\frac{\tau}{\alpha}}\right)}{\alpha - \beta},$$
$$\Phi(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}} - \alpha\beta c\left(e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}\right)}{\alpha - \beta},$$

where the new parameter c corresponds to $c = xk_2^+$ and the mean first passage time is

$$\langle\tau\rangle = \alpha + \beta - \alpha\beta c\,.$$

As expected, we find a much better agreement of the cumulative distribution function with randomly placed moving protein with a fit to the sequential two step process with mixed initial condition (see fig. D.10b) than with a single exponential distribution.

When the binding site on the polymer is placed initially at the surface such that it can be reached by the protein, we again observe a single exponential distribution when the protein diffuses so fast that it dominates the search process (see fig. D.11). Since only the protein is moving the process is again dominated by a single transition rate and can be described by a two state process.

With self-exclusion of the polymer we expect the search time distribution to show heavy tails when the polymer density is high. This is because some of the conformations of the densely packed self-avoiding polymer allow for very few movement, therefore leading to extremely large search times, much larger than expected from an exponential distribution. Due to the massive slowdown of the simulations in this regime, we can only observe the onset of the formation of heavy tails (see fig. D.12), whereas higher polymer densities are not feasible anymore.

3 Discussion

Kinetic Monte Carlo simulations have been used to characterise the target search kinetics of a polymer in confinement and a protein at the surface of the confining volume. Different strategies of speeding up the search process can be identified depending on the polymer density in the volume.

In accordance with asymptotic calculations of the MFPT for non-Markovian processes [169] we observe a linear dependence of the search time on the polymer length at lowest polymer densities, where the search is dominated by the centre of mass diffusion of the polymer. At intermediate densities the subdiffusive motion of the binding site leads to a compact search and the search time increases only with the square root of the polymer length, as predicted by non-Markovian theory.

While at low polymer densities the size of the polymer plays a major role, the reflecting boundaries lead to a saturation of the search time when polymer density becomes high, in accordance with previous simulations [85]. This is an important effect contradicting our intuition of the search process being complicated by the enormous length of e.g. a bacterial or eukaryotic chromosome. When an ideal polymer is long enough to frequently encounter the boundary, polymer segments become uncorrelated in their motion. Therefore the search time saturates at the value for a polymer that has the length of the correlated segment around the binding site. For a self-avoiding polymer only a short search time plateau is reached. Then the increasing polymer length leads to a slowdown of polymer motion and the mean search time diverges.

The effect of non-specific binding on the search process also depends on the polymer density in the volume. Without sliding of the protein non-specific binding at very high polymer densities slows down the search process because the binding site is very likely uncorrelated in its motion from the bound polymer bead and therefore not brought closer to the protein by non-specific binding. For lower polymer densities however an optimal relative binding rate can be identified, which increases as the polymer density drops further. Transient binding tethers the polymer to the surface and since the polymer is small compared to the volume, it brings the binding site closer to the protein. The lower the polymer density the more important becomes this effect and the optimal binding rate is replaced by a lower plateau when non-specific binding becomes so effective that even its negative effect of hindering protein diffusion becomes unimportant.

A similar effect of non-specific binding is found in dependence on the protein diffusion rate. While fast diffusing proteins are very efficient in their search and only hindered by non-specific binding, an optimal relative binding rate is found as the protein diffusion decreases to intermediate values. This optimal relative binding rate increases as the protein slows down further, since hindering protein diffusion becomes less important and the effect of bringing the polymer closer to the protein dominates.

In contrast to the target search of cytoplasmic transcription factors, where facilitated diffusion, i.e. the interplay of 1d sliding and 3d diffusion of the protein was shown to be an important tool to speed up the process, sliding of the protein along the polymer does not play a central role in our case of surface-located proteins. Since the possible sliding range is greatly limited by the polymer segments that are at the surface, even infinitely fast sliding speeds up the process by only a factor smaller than two. An optimal sliding rate is only identified for approximately permanent binding of the protein, because a large fraction of always the same polymer segments is tethered to the surface for very high sliding rates, therefore inhibiting polymer diffusion. When the binding site is likely to be close to the bound segments, i.e. for small volumes or small polymers, this does not hinder the search process and the minimum disappears.

When a static protein is placed at the corner of a cubic box, we observe dramatically larger search times than placing it at a face centre. This effect grows with increasing polymer length because the polymer is increasingly excluded from the corners. A polymer at a corner has a reduced number of accessible conformations, leading to a reduction of the conformational entropy and an increase in the free energy. In the limit of large volumes the change in free energy can be estimated by the free energy difference of a free polymer and a polymer with a

monomer tethered to the corner of an infinitely large cube. The estimated exponents for the scaling with polymer length are in partially good agreement with our simulations. Larger deviations are observed for a mid monomer reaching the face centre of the cube, however previous estimates of the exponent are only available for the mid monomer reaching the surface of a sphere, which might cause the observed deviations.

Cumulative distribution functions of the target search time have been analysed to study the underlying stochastic process of the search. When either protein diffusion or polymer movement dominates the search, good agreement with a single exponential distribution is found. This is the expected distribution for a Brownian particle reaching a target [174] and has also been reported for polymers searching for a static target in confinement [85]. When an approximately static protein is placed randomly at the surface, exclusion from the corners leads the cumulative distribution function to agree with a simultaneous multi step process, i.e. an overlay of multiple exponential processes with different rates depending on the initial condition (the location of the protein). When both protein diffusion and polymer motion are limiting for the process, we find a good agreement with a reversible sequential two step process. The intermediate state of this process could correspond to configurations where the binding site is already at the surface but not yet bound to the protein.

While the copy number of membrane-integrated transcription factors is typically tightly controlled and low (for CadC in *E. coli* on average 1–3 molecules per cell [81]), the number of nuclear pore complexes per nucleus was found to vary between 1–621 [175, 176], with varying density among cell types [175]. As we observe an inverse proportionality between the number of proteins and the mean search time for low protein densities, increasing the number of proteins could be a way to speed up the process of mRNA molecules locating nuclear pores to exit the nucleus. Moreover, for small polymers like mRNA molecules, our simulations predict that transient binding to the protein could facilitate the search. This is in accordance with the gene gating hypothesis [33], where active genes are brought next to NPCs, therefore bringing the mRNA close to the protein. As the search is much faster than for longer polymers and the binding sites are at the ends of the polymer, which are more mobile than other monomers, the process could also be fast enough without this additional speedup.

Taken together, our simulations reveal that elaborate strategies to speed up the target search process, like transient tethering of the polymer to the surface, are mainly relevant at low polymer densities, while at high densities a decorrelation of polymer subchains due to the confinement facilitates the search process even for very long polymers.

Search Dynamics of a Membrane-Integrated One-Component Receptor

IV

After the extensive study of the numerical target search simulations in the previous chapter, the application to a specific one-component receptor in *E. coli* will be discussed. The following chapter is adapted from the manuscript published in [177] and contains results obtained in collaboration with Sophie Brameyer and Elisabeth Hoyer from the group of Prof. Dr. Kirsten Jung, who carried out all experiments.

1 Introduction

In order to adapt to fluctuating environments with frequent changes in nutrient conditions and communication signals, but also life-threatening conditions such as environmental stresses and antibiotics [178], bacteria have evolved sophisticated signalling frameworks, primarily based on one- and two-component systems [18, 19]. While the target search dynamics of two-component systems, featuring a sensor kinase and a separate cytoplasmic transcription factor have been thoroughly studied, as discussed in chapter II, the kinetics of one-component systems with a membrane-integrated sensor that directly binds to the chromosomal DNA still harbour many open questions.

In one-component signalling systems, a single protein functions both as a sensor of environmental conditions and as a response regulator [18]. For those of them that are membrane-integrated sensors and DNA-binding transcription factors the open question is how they locate their specific binding site on the chromosomal DNA from the membrane. This is the case for one-component systems of the ToxR receptor family, for example the pH stress-sensing receptor CadC in *E. coli* [24].

CadC is a particularly well studied membrane-integrated one-component receptor. It is part of the Cad system in *E. coli*, which is a pH stress response system that also depends on signalling input from the lysine-specific permease LysP [82]. The Cad system alleviates acidic stress by activating the synthesis of the lysine/cadaverine antiporter CadB and the lysine decarboxylase CadA. CadA catabolises the decarboxylation of lysine to cadaverine and CO_2, which counteracts acid stress through consumption of cytosolic protons. Cadaverine is then exported by CadB in exchange for lysine. CadA and CadB are encoded by the *cadBA* operon, which is transcriptionally upregulated by specific binding of CadC upstream of the promoter [179], shown in fig. IV.1. Activation of CadC is caused by an acidic pH in the presence of external lysine and inhibited by the presence of cadaverine. The transient response of CadC leads to tightly controlled *cadBA* transcription and is required because under stress conditions

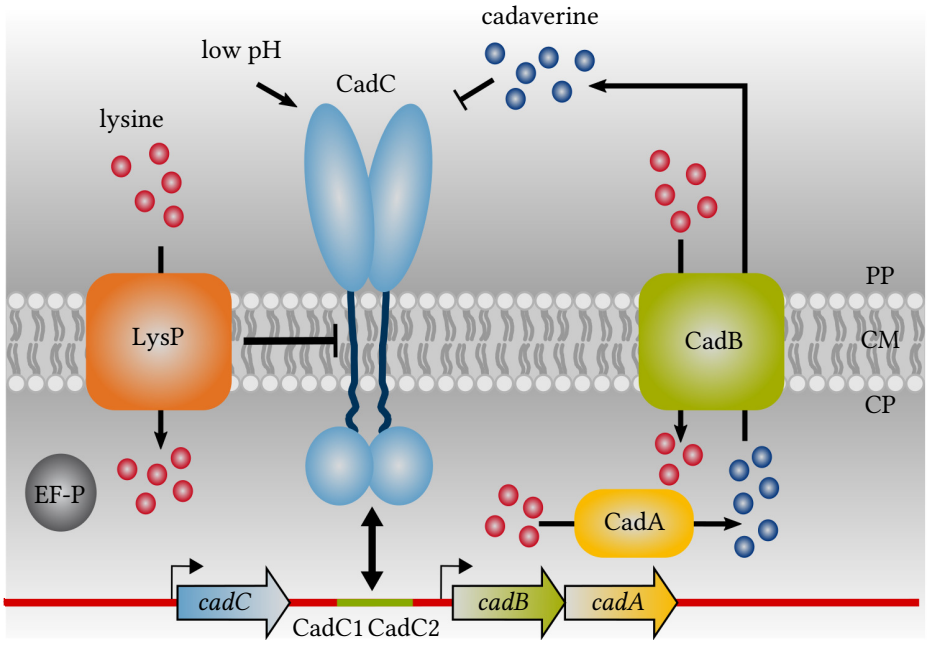


Figure IV.1: Cad system in *E. coli*. The Cad system is a pH stress response system in *E. coli*. The one-component receptor CadC (blue) is a membrane protein with a periplasmic (PP) domain, a transmembrane (CM) helix and a cytoplasmic (CP) DNA-binding domain. It is activated by acidic pH in the presence of external lysine (red) and inhibited by the presence of cadaverine (dark blue). Upon activation CadC homodimerises and binds to the two CadC binding sites within the *cadBA* promoter, inducing the expression of CadA and CadB. The lysine decarboxylase CadA (yellow) converts lysine to cadaverine, which is exported by the lysine/cadaverine antiporter CadB (green). While the external pH and the presence of cadaverine are both sensed by the periplasmic domain of CadC, the availability of external lysine is transduced to CadC via the co-sensor and inhibitor LysP (orange) by inhibiting homodimerisation of CadC. The elongation factor P (EF-P) is shown in grey.

CadA quickly increases to up to 2 % of cellular protein [83]. Both sensing changes in the external pH and binding to cadaverine was found to proceed within the periplasmic domain of CadC [180, 83], whereas the availability of external lysine is transduced to CadC via the co-sensor and inhibitor LysP by inhibiting homodimerisation of CadC [181, 182]. A drop in external pH induces dimerisation of the periplasmic sensory domain of CadC, leading to a structural rearrangement of the cytoplasmic linker and homodimerisation of the DNA-binding domain [180, 183, 24]. The CadC protein number is extremely low (on average 1–3 molecules per cell [81]), mainly due to a low translation rate caused by polyproline stalling, which is only partially relieved by elongation factor P (EF-P) [184].

The three models of how membrane proteins can be localised in bacterial cells, described earlier and depicted in fig. I.5, have been experimentally evaluated for CadC, leading to the best agreement with the diffusion and capture mechanism [81, 82, 83]. An observation arguing against proteolytic cleavage is that extracellular cadaverine rapidly deactivates the

CadC response after the original stimulus, which would not be feasible if the DNA-binding domain was separated from the rest of the molecule [82]. As for the transertion mechanism, moving the *cadC* gene to a locus far from its native position, which is close to the CadC target site on the DNA would be expected to reduce CadC activity. However, relocating *cadC* to the *lac* operon did not show any effect. The diffusion and capture mechanism however appeared consistent with experiments that imaged fluorophore-labelled CadC *in vivo* [81]. The experiments showed that after cells were shifted to a medium providing acid stress and a lysine-rich environment, localised CadC spots appeared in fluorescent microscopic images and disappeared again upon removal of the input signals. Moreover, the number of CadC spots was positively correlated with the number of DNA binding sites, indicating that the spots correspond to CadC-DNA complexes with a much lower mobility than freely diffusing CadC in the membrane.

Therefore, the existing data suggest that the one-component system CadC establishes the protein-DNA contact required for transcription regulation not by a conventional target search akin to cytoplasmic transcription factors, but instead by 2d diffusion of the protein in the membrane and fluctuations of the DNA conformation that occasionally bring the DNA region of the target site close enough to the membrane to be captured by the protein. Intuitively, a successful diffusion and capture event seems highly unlikely. However, that such events occur was independently demonstrated in an experiment that artificially tethered the Lac repressor to the cell membrane [185], which was still able to inducibly repress transcription from a chromosomal reporter. Hence, the striking questions are how this type of target search is kinetically feasible and on which time scale.

Here, we address these questions using *E. coli* CadC as a model system. We measure the search time of CadC to its target DNA binding site in single cells by probing the formation of fluorescent CadC spots at different times after a medium shift to low pH and rich lysine conditions. This yields experimental search time distributions that we compare against first passage time distributions obtained from stochastic models [186]. We also measure the mobility of a chromosomal locus in our experimental setup using a fluorescent repressor/operator system to inform our kinetic Monte Carlo simulations of the target search dynamics. These simulations are able to reproduce the experimental behaviour and to elucidate properties of the search process that we cannot obtain experimentally.

2 Results

2.1 Experimental Search Time Distribution

Before summarising the main results of investigating the target search of CadC for its specific binding site on the DNA, we discuss the experimental measurements and data analysis leading to search time distributions that are comparable to our simulations.

2.1.1 Experimental model system

In the present study, CadC serves as an experimental model system to investigate the DNA target search of a membrane-integrated transcription factor, depicted in fig. IV.2. To probe

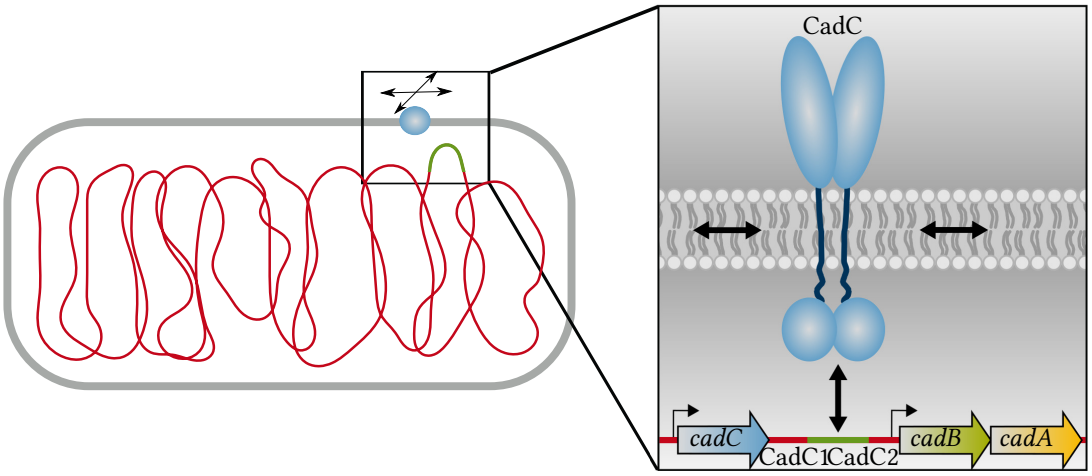


Figure IV.2: Target search by CadC. The target search of membrane-integrated transcription factors is investigated by experimentally measuring the response time of CadC. The molecular model on the right shows CadC in blue, diffusing in the membrane and forming two dimers to bind to the two CadC binding sites within the *cadBA* promoter (CadC1 and CadC2), displayed in green. The *cadC* gene is located upstream of the *cadBA* operon.

the kinetics of this search process in individual cells, it is crucial to have a well-defined initial state and clear "start" and "stop" events. When cells are initially grown in a medium with neutral pH, CadC is inactive and homogeneously distributed in the membrane [81]. A sudden medium shift to low pH and rich lysine conditions then serves as a suitable "start" trigger, which activates CadC via the signal transduction mechanism described above and starts the target search process. Dimerisation of CadC is expected to occur on a much faster time scale than the protein-DNA search process, as discussed below. Detecting the successful termination of the target search is challenging. Here, we exploit the previous finding that the formation of stable CadC-DNA complexes is visible as distinct spots in fluorescence microscopy images [81]. The study excluded that spot formation occurs solely due to low pH, using a pH-independent CadC variant that showed spots at both neutral and low pH. The connection between spot formation and CadC DNA-binding was derived from the observation that no spots were formed when CadC was rendered unable to bind DNA, and the number of spots per cell correlated with the number of DNA binding sites for CadC. When the DNA binding site was deleted, only 20 % of cells formed spots, possibly due to non-specific binding to the DNA. We take this into account in our quantitative analysis below.

2.1.2 Experimental measurement of CadC target search times

Three *E. coli* strains were used with different binding site configurations on the chromosome (see appendix E1): (i) N-P_{cadBA} (wild type), with the native DNA binding site at 93.9′, relatively close to the origin of replication (*ori*) [187], (ii) T-P_{cadBA}, with the binding site relocated to the terminus, and (iii) N+T-P_{cadBA}, with both binding sites, illustrated in fig. IV.3. To visualise the temporal and spatial localisation of CadC *in vivo*, each of the strains was transformed

with plasmid-encoded mCherry-tagged CadC, which slightly increases the average number of CadC molecules per cell to 3–5 [81]. After the medium shift at $t = 0$ min, fluorescence and phase contrast microscopy images of different cells sampled from the same culture every minute were taken (see appendix E2).

Image analysis tools were used to detect fluorescent spots within the cells, evaluating between 859 and 2506 cells per time step. To analyse the fluorescence microscopy images for CadC or ParB spots within the cells, we used Oufti [188], an open-source software designed for the analysis of microscopy data for cell segmentation of the phase contrast microscopy images. The resulting cell outlines were used in a custom-written software implemented in Matlab to detect fluorescent spots. Briefly, a graphical user interface (GUI) was implemented that allows testing the parameters in a test mode before running the actual detection (see fig. E.3). In detection mode a function *SpotDetection.m* is called, that iterates through all frames and all cells. For each cell, from pixels in the fluorescence microscopy images the intensity of which is above a threshold defined by the parameters and dependent on the mean and variance of the fluorescence signal within the cell the connected components are computed. The components are checked for minimum and maximum size and minimum distance to other spots before being added to the list of spots. For further computations, information on all cells and spots were saved for all frames corresponding to a certain time after receptor activation.

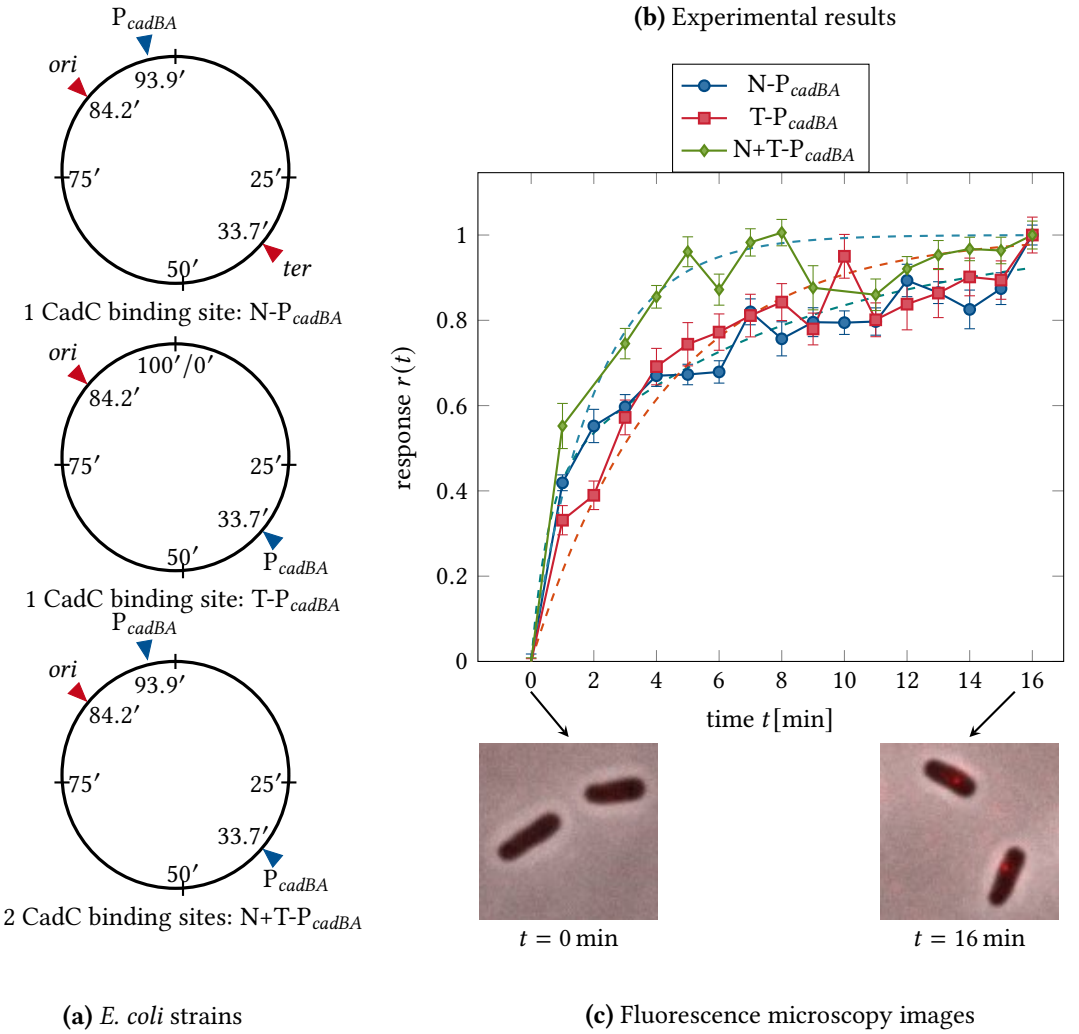
Based on these data, we determined the fraction of cells with at least one fluorescent spot, $\nu(t) = N_{\text{cells with spot}}(t)/N_{\text{cells}}(t)$ at each time t after CadC activation. To take into account the initial fraction of spots attributed to non-specific DNA binding (see above), we defined the response function

$$r(t) = \frac{\nu(t) - \nu(0)}{\nu(\infty) - \nu(0)}, \tag{IV.1}$$

which rises from zero to one. Here, the asymptotic value $\nu(\infty)$ accounts for the fact that fluorescent spots are never detected in all cells (see the raw data in fig. E.1). This is likely due to the heterogeneous distribution of CadC [81]: Given the low average copy number, some cells are expected to have less than the two molecules required for dimerisation. Additionally, some spots may have been missed by the spot detection algorithm, in particular for cells that were not perfectly in focus. The time-dependent response $r(t)$ for our three strains is shown in fig. IV.3, with examples of fluorescence microscopy images of cells at $t = 0$ min and the last time point. Before analysing the experimental response functions, we discuss the description of the target search as a stochastic process and derive theoretical response functions for comparison with the experimental data.

2.1.3 The target search as a stochastic process

To conceptualise the CadC target search process and the experimental response function, eq. (IV.1), we turn to a coarse-grained model, in which the CadC search for a specific binding site on the DNA can be described by a stochastic process with a small number of discrete states. Since both CadC and the DNA must move in order to establish a specific protein-DNA



(a) *E. coli* strains

(c) Fluorescence microscopy images

Figure IV.3: Experimentally measured target search by CadC. (**a**) Three *E. coli* strains with different positions of the *cadBA* promoter: N-P$_{cadBA}$ with the native DNA binding site close to *ori*, T-P$_{cadBA}$ with the binding site at the terminus and N+T-P$_{cadBA}$ with both binding sites. (**b**) Experimental results from CadC spot detection. Fluorescent microscopic images were taken every minute after receptor activation and analysed for CadC spots for the strains defined in panel B. The plot shows the response $r(t)$, defined as the normalised fraction of cells with fluorescent spots as a function of time t after exposure to acid stress. Error bars correspond to the propagated standard deviation of $v(t)$ from averaging over multiple data sets. The dashed lines in the plot show a fit of the response function to the CDF of a sequential reversible two-step model with mixed initial condition for N-P$_{cadBA}$, shown in blue dots and with fixed initial condition for T-P$_{cadBA}$ (red squares) and N+T-P$_{cadBA}$ (green diamonds). (**c**) The fluorescence microscopy images demonstrate how fluorescent spots appear after receptor activation.

contact, it is reasonable to assume a reversible sequential process with an intermediate state,

$$S_1 \underset{k_2^-}{\overset{k_1^+}{\rightleftharpoons}} S_2 \xrightarrow{k_2^+} S_3\,.$$

Here, state S_1 corresponds to configurations where the DNA target site is not in direct vicinity of the membrane, while CadC is delocalised on the membrane, unbound to the DNA. State S_3 corresponds to the final state where CadC is bound to a specific target site on the DNA. The intermediate state S_2 could then correspond to configurations where the DNA segment containing the target site is close to the membrane, but CadC is not bound to this segment. The transition rates between these states are denoted as k_1^+, k_2^+, and k_2^-.

We are interested in the first passage time τ to reach the final state S_3, which corresponds to the target search time within this coarse-grained description. The probability distribution $\Psi(\tau)$ for this time is calculated by making the final state absorbing, using standard techniques [186] (see appendix C1). Assuming that the system is initially in state S_1, the first passage time distribution is

$$\Psi(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}}{\alpha - \beta}\,, \tag{IV.2}$$

where the two time scales α and β of the exponential functions are related to the transition rates via

$$\alpha = 2\left(k_1^+ + k_2^+ + k_2^- - \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+k_2^+}\right)^{-1}\,,$$
$$\beta = 2\left(k_1^+ + k_2^+ + k_2^- + \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+k_2^+}\right)^{-1}\,,$$

implying that $\alpha > \beta$. At large times, the distribution $\Psi(\tau)$ decays exponentially (decay time α), whereas the time scale β corresponds to a delay at short times. Hence, increasing α leads to a slower decay and increasing β to a longer delay.

To relate the first passage time distribution to our experimental response function, eq. (IV.1), we consider the cumulative distribution function (CDF) defined by

$$\Phi(\tau) = \int_0^\tau \Psi(t)\mathrm{d}t\,,$$

which is the probability that the first passage time is less than or equal to τ. Experimentally, $\Phi(\tau)$ corresponds to the fraction of cells in which the target search was successful by time τ. The response function in eq. (IV.1) is our best proxy for this fraction of cells, and hence we identify

$$\Phi(\tau) \mathrel{\widehat{=}} r(\tau)\,,$$

such that one can use the CDF of the reversible sequential process as a fitting function for our data. The cumulative distribution function for $\Psi(\tau)$ of eq. (IV.2) is

$$\Phi(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}}}{\alpha - \beta}\,. \tag{IV.3}$$

However, using eq. (IV.3) amounts to the assumption that all cells are initially in state S_1. If we allow for the possibility that some cells are in state S_2 when CadC is activated, we have a mixed initial condition, where the process starts either from state S_1 or from state S_2. Denoting the fraction of cells that are initially in state S_2 by x, such that a fraction $1-x$ starts in state S_1, we obtain a first passage time distribution and associated cumulative distribution function of the form

$$\begin{aligned}\Psi(\tau) &= \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}} + c\left(\alpha e^{-\frac{\tau}{\beta}} - \beta e^{-\frac{\tau}{\alpha}}\right)}{\alpha - \beta}, \\ \Phi(\tau) &= 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}} - \alpha\beta c\left(e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}\right)}{\alpha - \beta},\end{aligned} \tag{IV.4}$$

where the new parameter c corresponds to $c = xk_2^+$.

We use eq. (IV.3) and eq. (IV.4) as fit functions to describe our experimental response functions. We obtain the average and the variance of the search times analytically from the fit parameters α, β and c, including the statistical errors for both of these quantities (table IV.1), using error propagation and the full statistical covariance matrix for the parameters, see tables E.2 and E.3 and appendix C1. For instance, the average value of the distribution eq. (IV.2), referred to as the mean first passage time, is

$$\langle\tau\rangle = \alpha + \beta = \left(1 + \frac{k_2^-}{k_2^+}\right)\frac{1}{k_1^+} + \frac{1}{k_2^+}.$$

This corresponds to the average time for the first step ($1/k_1^+$) multiplied by the average number of trials needed to reach the second step, plus the average time for the second reaction step ($1/k_2^+$). In cases where one of the reaction steps is rate limiting, the mean first passage time is simply the inverse of the limiting rate, and the process simplifies to a two-state process with an exponentially distributed first passage time. Similarly, for eq. (IV.4), the mean first passage time is

$$\langle\tau\rangle = \alpha + \beta - \alpha\beta c\,. \tag{IV.5}$$

2.2 Mean Search Time

After computing the response function $r(t)$ for all three strains, we fit the data to the CDF of the theoretical models described above using the *curve_fit* function of the scipy module in Python, choosing a trust region reflective algorithm, which is an evolution of the Levenberg-Marquardt method that can handle bounds.

2.2.1 The mean search time is less than 5 min and not affected by relocation of the target site

We first analysed the experimental response $r(t)$ in the wild type strain. The wild type data (N-P$_{cadBA}$) in fig. IV.3 (blue dots) show an initially fast increase, followed by a more gradual saturation. This behaviour is not well described by eq. (IV.3), but captured by eq. (IV.4) with

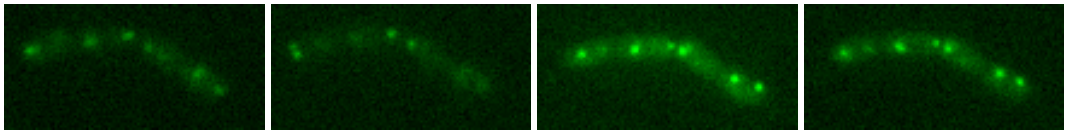


Figure IV.4: Fluorescent spots of ParB. A sequence of fluorescence microscopy images of the same cells showing ParB as fluorescent spots. The position of *ori* at low pH is estimated by inserting a *parS* gene close to it and letting ParB-yGFP bind to it.

the mixed initial condition, as can be seen by the fit to the data represented by the blue dashed line in fig. IV.3. From the fit parameters α, β, and c, we computed the mean search time according to eq. (IV.5), finding $\langle\tau\rangle \approx 4.84 \pm 0.19$ min. This result is consistent with the transcriptional response of the target genes *cadBA*, which was previously probed by Northern blot analysis, finding that the cell-averaged *cadBA* mRNA level starts to increase about 5 min after receptor activation [82].

In order to assess whether the position of the DNA binding site along the chromosome affects the target search, we analysed the behaviour of strain T-P$_{cadBA}$, which has the CadC DNA binding site at the terminus instead of the native position. The corresponding response function shown in fig. IV.3 (red squares) features a less pronounced initial increase than observed for N-P$_{cadBA}$, and the dashed red line shows an adequate fit using the sequential model with fixed initial condition in eq. (IV.3). The calculated mean first passage time of $\langle\tau\rangle \approx 4.20 \pm 0.15$ min is comparable to that for the wild type strain. The slower initial increase is compensated by a faster increase at later times to yield a slightly smaller mean search time. To characterise the shape of the mean first passage time distributions, we calculated the variance σ^2 of $\Psi(\tau)$, see table IV.1, which is smaller for strain T-P$_{cadBA}$ than for the wild type.

2.2.2 Search time is decreased with two chromosomal CadC binding sites

After observing essentially the same mean search time for two very distant locations of CadC target sites on the chromosome, we wondered how a strain harbouring both target sites would behave. We therefore repeated the measurements for *E. coli* strain N+T-P$_{cadBA}$, which has the native DNA binding site and additionally the binding site at the terminus. As shown in fig. IV.3 (green diamonds), the response function of this strain saturates much earlier than

Strain	$\langle\tau\rangle$ **[min]**	σ^2 **[min^2]**
N-P$_{cadBA}$	4.84 ± 0.19	49.6 ± 8.8
T-P$_{cadBA}$	4.20 ± 0.15	17.6 ± 2.0
N+T-P$_{cadBA}$	2.02 ± 0.12	4.09 ± 0.66

Table IV.1: Means and variances of the search time distributions. Results from fitting the experimentally computed CDFs to the sequential reversible model with mixed initial condition (N-P$_{cadBA}$) and fixed initial condition (T-P$_{cadBA}$ and N+T-P$_{cadBA}$). The fit parameters α, β and c were used to compute the mean first passage time $\langle\tau\rangle$ and the variance σ^2 with statistical errors obtained from error propagation using the full covariance matrix, see tables E.2 and E.3 and appendix C1.

for the other two strains. Fitting the response data to the sequential model with fixed initial condition (eq. (IV.3)), we obtained a mean first passage time of $\langle\tau\rangle \approx 2.02 \pm 0.12$ min, which is only around half of the time than for a single chromosomal binding site.

2.2.3 Colocalisation of CadC spots with the DNA binding site

We also wondered whether the fluorescence spots indicating the position of stable CadC-DNA complexes in single cells would show a similar spatial distribution as the *cadBA* locus. We therefore analysed the localisation of CadC spots along the long axis of the cell in *E. coli* wild type. As an estimate for the position of the *cadBA* locus along the cell, we tracked the position of *ori* at low pH. Towards this end we inserted a *parS* gene close to *ori* and let ParB-yGFP bind to it, making the *ori* region visible as a fluorescent spot in microscopy images, demonstrated in fig. IV.4. The results from the image analysis were used to compute trajectories of ParB spots in a custom-written Matlab script by selecting the closest spots in subsequent image frames.

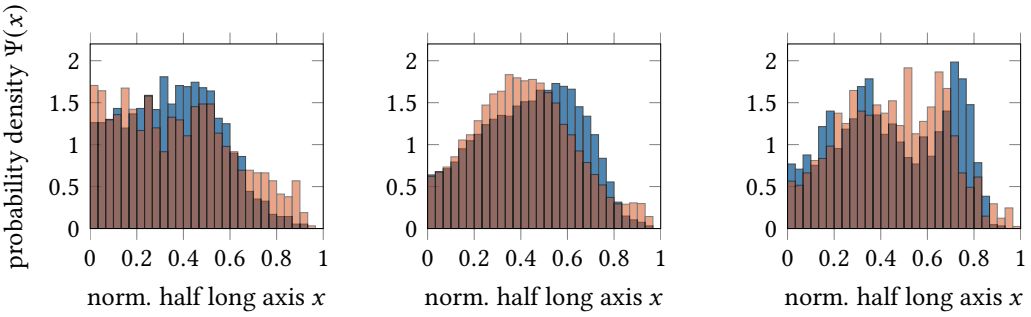
As the position of chromosomal loci depends on the progression of the cell cycle [6], we grouped cells according to their length into three classes. Figure IV.5 shows the spatial distribution of the relative spot positions along the half long axis for these three length classes, comparing ParB spots (blue) to CadC spots (orange). The large overlap of the distributions implies a similar cell age dependent localisation of *ori* and CadC spots along the long cell axis, suggesting that CadC spots indeed form close to the DNA binding site.

2.3 Numerical Simulations

To gain more insight into the dynamics of the target search, we turned to a coarse-grained biophysical model for the coupled dynamics of CadC and the DNA. We simulated the search of a CadC dimer for its target DNA binding site using a lattice model and a kinetic Monte Carlo approach.

2.3.1 Biophysical model

Since the target search of CadC for its DNA binding site(s) has been experimentally shown to succeed within a few minutes, we were aiming at finding the simplest coarse-grained model that can rationalise this fast response. We constrained our simulations to the main components of the target search: DNA with one or two specific binding sites moving inside an *E. coli* cell represented by a cubic lattice and one CadC bead diffusing in the cell membrane, depicted in fig. IV.6. Starting from a random initial configuration, the DNA molecule moves in the cytoplasm and the CadC dimer diffuses in the membrane until it reaches the specific DNA binding site to end the search process. This reduced model does not account for non-specific binding and sliding of the dimer on the DNA. As we take the average over different initial conditions, the small fraction of initial states that are already close to the target state and for which the search time is overestimated by the approximation should be negligible given our typical lattice sizes.



(a) Cell length (1 to 2)Δl **(b)** Cell length (2 to 6)Δl **(c)** Cell length (6 to 10)Δl

Figure IV.5: Localisation of *ori* and CadC spots. Localisation probability of CadC spots in N-P$_{cadBA}$ cells (orange) and ParB localisation marking *ori* (blue) along the half long axis of cells. The half long axis is normalised such that mid-cell is at $x = 0$ and the poles are at $x = 1$. Overlaps of the two distributions are shown in darker orange. Cell age is taken into account by splitting all occurring cell lengths into ten equally spaced steps Δl and pooling the cells according to their size. From the ten different age classes we observed similar localisation probabilities for $l = (1 \text{ to } 2)\Delta l$ (**a**), $l = (2 \text{ to } 6)\Delta l$ (**b**) and $l = (6 \text{ to } 10)\Delta l$ (**c**), which are therefore grouped together in this plot.

2.3.2 CadC dimerisation

CadC dimerisation was not simulated explicitly, instead the CadC beads on the cell surface correspond to already formed dimers. To justify this, we estimate the time of dimerisation for two CadC monomers using the approximate formula eq. (I.2) for the mean time for a particle moving with diffusion constant D on the surface of a sphere of radius R to first encounter a trap of radius a_t. To estimate the time of dimerisation for two moving CadC monomers with diffusion constant D_1 and radius r_1, we insert the relative diffusion constant $D = 2D_1$ and the contact radius $a_t = 2r_1$. Using a sphere of the same typical surface area as a cell, 4.40 µm^2 [189], we obtain $R = 0.59$ µm. For the radius of a CadC monomer we use $r_1 = 2.40$ nm [190] and for the CadC monomer diffusion constant $D_1 = 0.34$ µm^2 s^{-1}, which is a typical value for a membrane protein with a single transmembrane domain [191]. We then obtain a mean first encounter time of 5.1s, much faster than the total observed response time of several minutes and therefore allowing us to omit this step in our biophysical model of the target search.

2.3.3 Realistic simulation parameters

To compare our target search simulations to our experiments we used parameter values based on experimental estimates, summarised in table IV.2. In the following, we refer to these parameter values as the "realistic parameter set". Since the move rates for CadC dimers and the DNA molecule are based on *in vivo* measurements, they implicitly take into account effects due to crowding in the cytoplasm and in the membrane.

The *E. coli* cell volume of $\propto 1$ µm^3 is approximated by a simulation box of volume L^3 with $L = 12a$ and lattice constant $a = 88.42$ nm, yielding a cell volume of $V \approx 1.20$ µm^3. The *E. coli* chromosome with 4 639 221 bp [192], which measures 1.58 mm corresponds to a

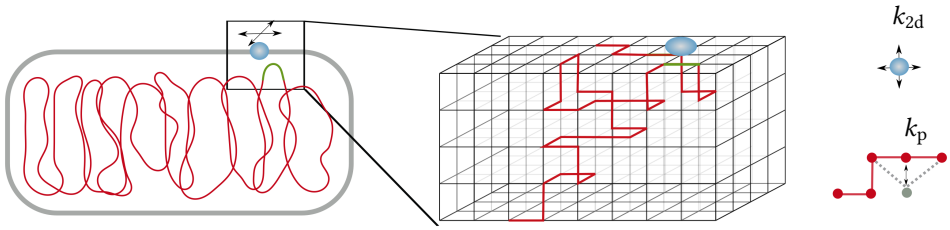


Figure IV.6: Simplified lattice model. The target search of membrane-integrated transcription factors is studied by kinetic Monte Carlo simulations of a lattice model. The movement of a CadC dimer (blue) is restricted to the surface of the lattice, whereas the segments of the DNA contour (red) can be anywhere within the lattice. Proteins diffuse on the surface with rate k_{2d}, the polymer moves by random displacement of single beads with rate k_p, without excluded volume interactions between beads or bonds.

length of 15773a. Compared to our discussion in the previous chapter we are therefore in a regime of very high polymer density. Since our simulations have shown that the target search dynamics are independent of polymer length once it has reached a cell size dependent critical length (see fig. III.2a), we save computation time by choosing the polymer length well above this threshold at 100a. It has been reported that the diffusion coefficient for protein diffusion within the cell membrane depends primarily on the size of the transmembrane domain [193, 194]. A CadC dimer has two transmembrane helices like the membrane protein WALP-KcsA, for which the diffusion constant was measured in multiple studies, yielding values between 0.21 $\mu m^2\ s^{-1}$ [194] and 0.25 $\mu m^2\ s^{-1}$ [195]. In a recent study the relevant CadC diffusion coefficient was also measured, albeit with a different fluorophore (mNG) than in the present study [81]. Those measurements yielded values ranging from 0.07 $\mu m^2\ s^{-1}$ to 0.19 $\mu m^2\ s^{-1}$. For our estimate of the model parameter values, we used the rounded value of $D_{2d} = 0.20\ \mu m^2\ s^{-1}$, which lies in the range of the quoted figures.

Parameter	**Symbol**	**Value**
lattice constant	a	88 nm
cell volume	V	1.2 μm^3
DNA length [beads]	N	100
number of CadC dimers	n_p	1
number of binding sites	n_b	1
CadC diffusion coeff.	D_{2d}	0.20 $\mu m^2\ s^{-1}$
DNA diffusion coeff.	D_0	0.0060 $\mu m^2\ s^{-1}$

Table IV.2: Simulation parameters matching the experiments.

2.3.4 Mobility measurements of the DNA molecule

To complete our parameter set, we estimated the diffusion constant D_0 of a DNA segment in our simulations by experimentally tracking a chromosomal locus. The origin was tagged using a *parS*/ParB fluorescent operator/repressor system (FROS) as discussed above. For N-P$_{cadBA}$-*parS_ori* and Δ*cadC*-*parS_ori*, a wild type and a mutant *E. coli* strain lacking *cadC*, each containing *parS_ori*, fluorescence and phase contrast microscopy time-lapse videos were taken of the same cells every 30 s for activating and inactivating conditions, respectively.

The results from the image analysis were used to compute trajectories of ParB spots in a custom-written Matlab script by selecting the closest spots in subsequent image frames. From the trajectories of ParB spots the ensemble averaged mean squared displacement (MSD) was computed as a function of time lag τ: $\mathrm{MSD}(\tau) = \langle(\vec{r}(t) - \vec{r}(t+\tau))^2\rangle$, where the mean was taken over different spots and $\tau = n30\,\mathrm{s}$ with $n \in \{1, \ldots, N\}$ and the number of time steps N. The resulting MSD curves (see fig. E.2) are in semi-quantitative agreement with other tracking experiments of chromosomal loci in *E. coli* [196, 197]. Previous experiments showed that DNA diffusion in *E. coli* agrees well with the Rouse model [105], with diffusion exponents in the range 0.40–0.60 [196, 198, 197]. The Rouse model predicts the MSD in 2d [199]:

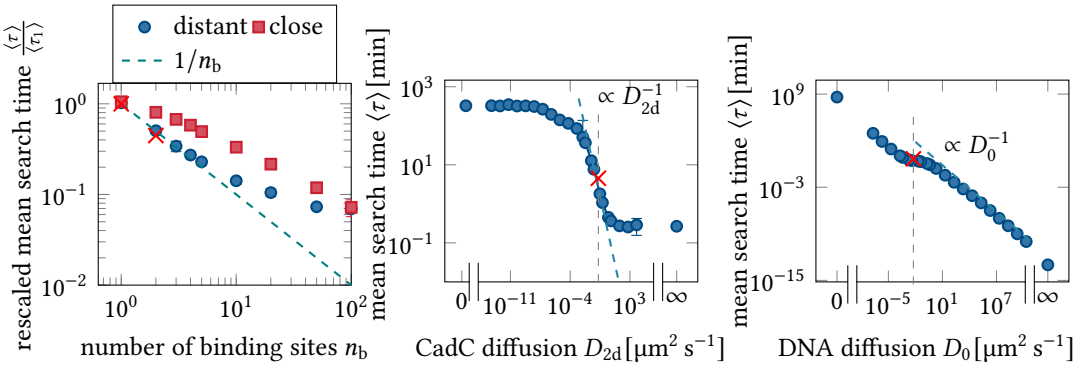
$$\begin{aligned}\mathrm{MSD}(\tau) &= \sqrt{\frac{16 l_\mathrm{K}^2 k_\mathrm{B} T}{3\pi\zeta}}\tau^{0.5} \\ &= \sqrt{\frac{16 l_\mathrm{K}^2}{3\pi}} D_0^{0.5}\tau^{0.5},\end{aligned}$$

with Boltzmann's constant k_B, absolute temperature T, Kuhn length l_K, friction constant ζ and bead diffusion constant D_0, where we have used $D_0 = k_\mathrm{B}T/\zeta$ [102]. We fitted the experimentally determined MSD to $\mathrm{MSD}(\tau) = \Gamma\tau^{0.5}$ and determined the diffusion constant from $D_0 = \Gamma^2 3\pi/(16 l_\mathrm{K}^2)$.

We obtained $\Gamma = 0.0111 \pm 0.0001\,\mu\mathrm{m}^2\,\mathrm{s}^{-0.5}$ for *E. coli* wild type under activating conditions and $\Gamma = 0.0091 \pm 0.0001\,\mu\mathrm{m}^2\,\mathrm{s}^{-0.5}$ for Δ*cadC* under inactivating conditions. The DNA mobility seems to be independent of the probed conditions, since the two values do not differ a lot. Hence we used the average value $\langle\Gamma\rangle$ to compute the diffusion constant D_0 of a Kuhn segment, yielding $D_0 \approx 0.0060\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, which is the value chosen for the realistic parameter set. To convert the dimensionless search times τ' from the simulations into seconds we computed $\tau' a^2/D_0 \approx 1.30\tau'$.

2.4 Simulation Results

After having defined a realistic parameter set, we discuss the main results of the numerical simulations of the reduced target search model and how they quantitatively compare to the experimental findings.



(a) Multiple binding sites **(b)** CadC diffusion **(c)** DNA diffusion

Figure IV.7: Characterisation of the target search by computer simulation. (**a**) Dependence of the mean search time $\langle\tau\rangle$, normalised by the search time with a single binding site $\langle\tau_1\rangle$, on the number of DNA binding sites n_b. Realistic parameters are used for binding sites placed either uniformly (blue dots) or next to each other (red squares) along the circular chromosome. Red crosses correspond to the experimental results for one (N-P$_{cadBA}$) and two (N+T-P$_{cadBA}$) DNA binding sites. (**b**) Mean target search time as a function of the CadC diffusion coefficient. The leftmost data point corresponds to $k_{2d} = 0$ and the rightmost data point corresponds to $k_{2d} \to \infty$, simulated by covering the whole cell surface by proteins. The estimated diffusion coefficient of a CadC dimer, $D_{2d} = 0.20\,\mu m^2\,s^{-1}$ is marked (grey dashed line), as well as the experimentally measured search time for N-P$_{cadBA}$ (red cross). (**c**) Mean target search time as a function of the DNA diffusion coefficient. The dashed grey line marks $D_0 = 0.0060\,\mu m^2\,s^{-1}$, used in the realistic parameter set, and the red cross marks the experimentally measured search time for N-P$_{cadBA}$.

2.4.1 Simulating multiple DNA binding sites

Motivated by the experiments with the N+T-P$_{cadBA}$ strain, we performed simulations with varying number of DNA binding sites. In fig. IV.7a the mean search time $\langle\tau\rangle$ normalised by the mean search time with a single DNA binding site $\langle\tau_1\rangle$ is shown as a function of the number of binding sites n_b. Placing the binding sites as far apart as possible on the circular chromosome, simulations with realistic parameters show a halving in search time when increasing the number of binding sites from one to two. This is the expected result for two binding sites moving independently of each other due to the decorrelation of polymer subchains in spatial confinement. How far two binding sites have to be apart along the DNA to behave as independent targets therefore depends on the size of the simulated cell. The initial inverse-n_b scaling of $\langle\tau\rangle$ flattens progressively as the binding sites come closer to one another and are more correlated in their movement. Placing the binding sites next to each other (red squares) has an expectedly smaller effect for small n_b, since it only increases the size of the binding site. The curve becomes steeper as the binding sites occupy a larger fraction of the polymer.

Given that the two DNA binding sites in N+T-P$_{cadBA}$ are on opposite sides of the chromosome, we expect them to behave as two independent binding sites. The experimentally

measured reduction of the search time by roughly a factor of two (red crosses in fig. IV.7a) is therefore in good agreement with our model. Also, by construction, the position of the binding site along the chromosome has no effect on the simulated search times.

2.4.2 Target search time is sensitive to CadC diffusion

To address the question whether the search process is more sensitive to changes in DNA mobility or CadC diffusion, we performed simulations with varying diffusion rates. Plotting the mean search time as a function of the CadC diffusion coefficient, D_{2d}, in fig. IV.7b, we observe three different regimes. For slow protein diffusion, the search time is almost independent of D_{2d}, followed by a range where the search process is entirely dominated by CadC diffusion ($\tau \propto D_{2d}^{-1}$), while for very fast CadC diffusion the search becomes less sensitive to D_{2d} again. The data point for infinitely fast CadC diffusion was obtained by making CadC cover the whole cell surface. The realistic value for the diffusion constant of CadC ($D_{2d} = 0.20\,\mu m^2\,s^{-1}$) is marked by a grey dashed line and lies in the regime where the search time strongly depends on D_{2d}. The red cross marks the experimentally measured mean search time for a single binding site, which is in surprisingly good agreement with the simulation data, given the simplicity of our biophysical model. Figure IV.7c shows the counterpart of fig. IV.7b for DNA diffusion, with $D_0 = 0.0060\,\mu m^2\,s^{-1}$ marked by a grey dashed line. While the mean search time is strongly dependent on DNA diffusion for small and large D_0, it is less sensitive to D_0 in the experimentally relevant intermediate regime.

To further analyse this observation, we used our simulations to approximate a target search exclusively due to CadC diffusion in the membrane. We used the realistic parameters but placed the DNA binding site at the membrane and set the DNA diffusion rate to zero, such that only CadC was moving, yielding a mean search time of $\langle\tau\rangle \approx 20\,s$. This is consistent with an estimate based on the previously reported [64] approximate formula eq. (I.2) for the mean first encounter time of a particle moving with diffusion constant D on the surface of a sphere with radius R with a trap of radius a_t. Approximating the cubic cell by a sphere of radius $R = 0.69\,\mu m$ and CadC with $a_t = 0.050\,\mu m$, yields $\tau \approx 13\,s$. This estimate also validates our simulations by showing that the dependence of the search time on the size of CadC is weak, rendering a correction for overestimating the size of CadC dimers in the simulations unnecessary.

While the DNA molecule is likely to be the less mobile part in the target search, it has to move at least close to the membrane to enable binding to CadC. The data point on the very right of fig. IV.7b corresponds to the time it takes the DNA binding site to bind anywhere to the membrane. It corresponds to $\tau \approx 16\,s$, similar to the time it takes CadC to locate the binding site at the membrane. Therefore, a scenario where the DNA binding site randomly reaches the membrane and CadC searches the membrane to bind to it seems to lead to reasonable search times according to our simulations.

For comparison we also simulated a target search process where CadC is immobile in the membrane and DNA diffusion has to account for the whole search. When only the DNA molecule is moving and all other rates are set to zero, with realistic values for DNA length and cell size a mean search time of $\langle\tau\rangle \approx 560\,min$ was calculated from the simulations, a response time that would not allow *E. coli* cells to survive the transition to acidic environments.

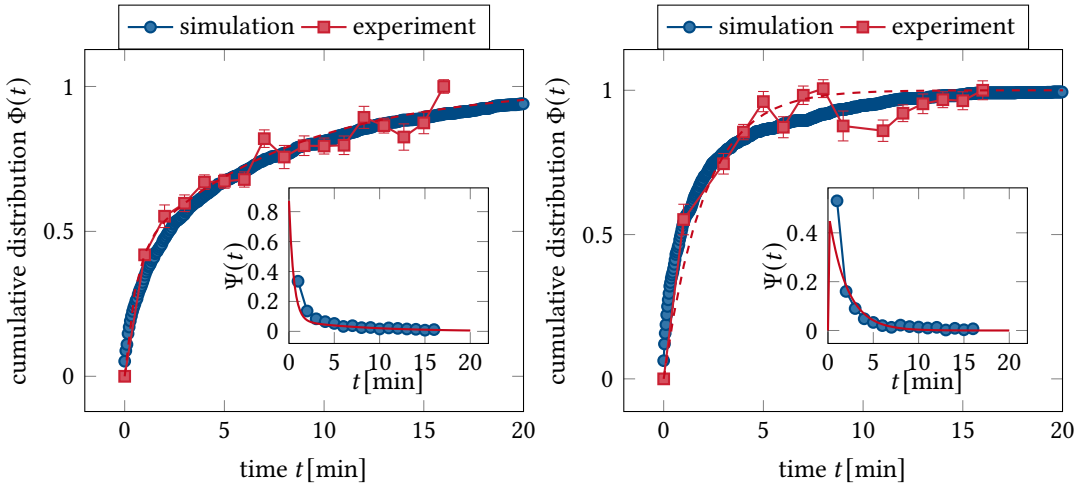
2.4.3 Quantitative comparison of simulated and experimental search time distributions

As shown above, the numerical simulations of the target search using our experimentally estimated parameter set yield mean search times that are fairly close to what we measured experimentally. To test whether the biophysical model can also quantitatively capture the experimental behaviour, we attempted to find simulations with time distributions closely matching the experimentally computed CDFs. In fig. IV.8a the experimental CDF of N-P$_{cadBA}$ together with the best fit to the sequential model with mixed initial condition is shown in red. In order to find a simulation with agreeing time distributions, we simulated the target search process using the estimated parameter set. As shown by the blue dots in fig. IV.8a this yields a very good accordance between experimental data and simulations. In fig. IV.8b the experimental CDF of N+T-P$_{cadBA}$ and the corresponding fit to the sequential model with fixed initial condition is shown in red. Our attempt to find a matching simulation by using the same parameters as for the simulations in fig. IV.8a but increasing the DNA binding sites to $n_\mathrm{b} = 2$ leads to quite a good agreement with the experiment, as shown by the blue dots. Despite the simplicity of the simulations, neglecting in particular constraints of DNA movement due to its specific organisation in the cell, the results agree surprisingly well with the experimental findings when using experimentally realistic parameters.

To further investigate the simulated first passage time distributions, we computed them for different parameter settings and fitted to the sequential model with fixed initial condition (eq. (IV.3)) and with mixed initial condition (eq. (IV.4)) respectively. While the shape of the first passage time distribution is independent of the system size parameters, the dynamic parameters have a big effect. For most parameter settings, including the realistic parameters, we found the best agreement of simulated CDFs with the sequential model with mixed initial condition. However, for simulations with an immobile polymer or slowly diffusing CadC, the search time distribution agrees rather with the sequential model with fixed initial condition. Since both in the experimental and simulated CDFs the delay β is very small compared to α, a fit to the fixed initial model gives virtually the same result as a single exponential CDF. When either movement of the DNA or of CadC accounts for most of the search there is only one limiting rate in the process, which can therefore be described by a single exponential distribution.

3 Discussion

We combined fluorescence microscopy experiments, quantitative analysis, and kinetic Monte Carlo simulations to characterise the target search kinetics of membrane-integrated transcription factors for a specific binding site on the chromosomal DNA. We were able to measure the time between the environmental stimulus and stable DNA-binding of a membrane-integrated one-component receptor, using the pH stress-sensing receptor CadC in *E. coli* as a model system. The measured mean search time of on average 4.5 min for a single DNA binding site is consistent with the time scale of the earliest transcriptional response [82]. Given the severe constraint of a membrane-anchored target search, it seems surprising that the search time is



(a) N-P$_{cadBA}$ **(b)** N+T-P$_{cadBA}$

Figure IV.8: Matching simulations to the experimental results. Numerical simulations are directly compared to the experimental distributions. Experimental data are shown in red squares, the red dashed line corresponds to the best fit. Blue dots correspond to the simulations that agree most with the experimental data. Insets depict the corresponding first passage time distributions. (**a**) The experimental CDF of N-P$_{cadBA}$ is shown together with the best fit and simulations using the realistic parameter set, summarised in table IV.2. A single binding site was placed at the centre of the polymer. (**b**) The experimental CDF of N+T-P$_{cadBA}$ is shown together with the best fit and simulations using the parameter set matching the experimental values with two binding sites on the polymer. The binding sites were placed far apart from each other and from the polymer ends. All simulations were started from a random initial configuration.

only about 5-fold slower than the search time of the cytosolic Lac repressor for its operator, which takes around one minute at a similarly low protein level [51].

As the position of the DNA binding site along the chromosome has no influence on the mean search time, the target search process appears to be quite robust. Given that the chromosome is highly organised within the cell, we believe this effect to arise mostly due to the mobility of CadC. While we do not have new evidence against proteolytic processing, this leads us to favour a diffusion and capture mechanism over transertion. Diffusion and capture mechanisms are well established for the localisation of membrane-integrated proteins like SpoIVB in *B. subtilis* [79, 80]. As the search time is independent of the position of the DNA binding site, transertion of CadC is at least not a requirement for fast response, in agreement with a previous evaluation of the three models [81]. Our finding that the mean search time decreases by a factor of two in a mutant with two DNA binding sites is also consistent with the diffusion and capture mechanism. Our simulations show that this is the expected result for two independent and equally accessible binding sites, where distant parts of the polymer become uncorrelated in their motion due to the confinement in the cell.

Despite the simplicity of our biophysical model of the target search, we simulated search times that match the experimental measurements surprisingly well. Our simulations with

a fixed protein number (one CadC dimer) produced search time distributions that were compatible with the experimental ones both in terms of their mean values and their width, suggesting that cell-to-cell heterogeneity in the number of CadC dimers does not significantly broaden the search time distribution. The constraint of chromosomal loci to macrodomains, which is neglected by our model, likely does not play a major role on the relatively short time scale of the search process, given that lateral motion of the CadC locus into the vicinity of the membrane is sufficient. Although we did not explicitly model crowding effects in the cytoplasm and at the membrane, the general slowdown of the kinetics due to crowding should be accounted for by our *in vivo* diffusion coefficients. Another open question is the lifetime of CadC dimers in the membrane. Active degradation should be negligible compared to the time scale of the search process. The majority of proteins in *E. coli* are not subject to rapid degradation [200] and individual membrane proteins that have been studied in detail displayed life-times as long as 6 days [201]. Spontaneous separation of CadC dimers from the membrane (or proteolytic cleavage) also do not appear to play a role, since CadC is deactivated by external cadaverine [82].

While it is difficult to experimentally answer the question whether the time scale of DNA motion or CadC diffusion limits the search process, the consistency between our simulations and experimental findings allows to draw some conclusions from an analysis of the simulations. In the experimentally relevant regime, the simulated mean search time is inversely proportional to the diffusion coefficient of the CadC dimer, but is relatively insensitive to the DNA diffusion speed. The search process therefore seems to be predominantly limited by the mobility of the transcription factors in the membrane. Since within the typical range of diffusion coefficients of membrane-integrated proteins [193, 194], the mean search time varies inversely proportional with the diffusion coefficient, this result is expected to hold for DNA-binding membrane proteins in general.

Beyond mean values the experimental data allowed the extraction of the cumulative distribution functions (CDF) for the target search time. Describing them by the distribution obtained from a reversible sequential two-step process resulted in a good fit for the wild type data assuming a mixed initial condition. The same was true for our simulations of the target search process with most parameter settings. Using a parameter set estimated for the experimental conditions led to a distribution that matched the experimental CDF of *E. coli* wild type. Attempting to find a matching simulation for the mutant with two DNA binding sites by using the same parameters with two DNA binding sites also yielded a good agreement. In contrast to the simulations, however, the experimental CDF of the mutant with two binding sites and the mutant with the binding site at *ter* show a better agreement with the sequential model with fixed initial condition (or a single-exponential CDF). How exactly the position of the DNA binding site affects the initial condition of the search process is not clear at this point. In the simulations, distributions agreeing with the sequential model with fixed initial conditions are obtained when either only CadC is moving, or it is moving very slowly.

Taken together, our experiments and simulations indicate that CadC is highly mobile in the membrane, while the *cadBA* promoter on the *E. coli* chromosome is mobile enough to randomly reach the membrane, enabling CadC to locate the DNA binding site within about five minutes, independent of its position along the chromosome. While diffusion and capture mechanisms are established for the polar localisation of membrane proteins [79, 80, 202],

our study indicates a broader relevance of diffusion and capture mechanisms for the largely uncharacterised interactions of membrane-integrated proteins with chromosomal DNA [71].

V Summary and Outlook

Despite the enormous diversity of life on earth, there are some principles that are common to all living systems. One of them is the interaction between polynucleotides and proteins, which is of vital importance to a vast number of cellular processes. Since biopolymers such as DNA and RNA are typically spatially constrained to a certain volume (cell, nucleus), this interplay often amounts to a protein moving on the surface of a volume having to locate a specific binding site on a polymer inside the confining volume. Despite its prevalence, the dynamics of this target search problem have remained largely uncharacterised.

We employed kinetic Monte Carlo simulations and mathematical analysis to study the search kinetics of a surface protein for a binding site on a polymer inside the confining volume. Simulating a polymer representing a chromosome in the eukaryotic nucleus or DNA in a bacterial cell for a time long enough to capture the target search dynamics is computationally challenging. We therefore started by establishing a coarse-grained lattice model that allowed us to efficiently probe the dynamics. Evaluating different cubic lattice models for the polymer dynamics, a phantom chain single-site Bond-Fluctuation model and a ssBFM with Fraenkel springs connecting the beads were found to reproduce stable dynamics in agreement with the Rouse model. Based on these two models a simulation framework was implemented that enables us to simulate the encounter dynamics of confined polymers with moving proteins at the surface of the confinement. Motivated by the ability of membrane-integrated response regulators to bind non-specifically to DNA, non-specific binding and unbinding as well as one-dimensional sliding of the protein along the polymer was included in our simulations.

Extensive analysis of the numerical target search simulations led us to identify various strategies of speeding up the procedure mainly depending on the polymer density in the volume. For dense systems, the reflecting boundaries result in subchains of the ideal polymer becoming uncorrelated in their motion. Therefore the mean search time saturates and becomes independent of polymer length, contradicting our intuition that the search is complicated by the enormous length of most biopolymers. Transient tethering of the polymer to the surface by non-specific binding of the protein only slows down the process when polymer density is high, since the bound polymer segment is likely to be uncorrelated from the cognate binding site and therefore the binding site is not brought closer to the protein.

In the opposite regime of lower polymer densities, transient tethering due to non-specific binding of the protein to the polymer emerges as a way to speed up the process, leading to an optimal relative binding rate. With decreasing polymer density the optimal binding rate increases, since the effect of bringing the polymer closer to the protein increasingly dominates the negative effect of hindering protein diffusion. Whether transient tethering

is favourable also depends on the surface diffusion of the protein. When the protein is very fast, transient binding leads to a deceleration of the search process. With decreasing protein diffusion however the effect of bringing the polymer closer to the protein starts to dominate and transient tethering gains importance.

In contrast to the facilitated diffusion model for cytoplasmic proteins, where sliding along the polymer was shown to play a central role to explain the experimentally measured fast response times, our simulations revealed a very limited effect of sliding for surface located proteins. Mainly because the sliding range is greatly limited to a few polymer segments at the surface, sliding rarely has an effect on the mean search time by more than a factor of two.

Studying the cumulative distribution functions of the target search time we found that the procedure can be well described by either a two state or a three state stochastic process, depending on the diffusion rate of protein and polymer. When either protein diffusion or polymer movement dominates the search, agreement with a single exponential distribution is found and the process is well described by a single rate. When both protein diffusion and polymer motion are limiting factors of the process, a sequential two step process is required to describe the procedure.

Since for high polymer densities the search dynamics have been shown to be dominated by protein diffusion on the surface and conformational changes of the polymer without the necessity of non-specific binding or sliding to speed up the procedure, the target search of the membrane-integrated transcription factor CadC in *E. coli* was simulated with a reduced lattice model. Despite the simplicity of the model, considering only the DNA molecule moving in the cell with either one or two specific binding sites and a single CadC dimer moving in the membrane, quantitative agreement with the predictions from fluorescence microscopy experiments was found. Using a parameter set based on previous measurements as well as our own computations to match the experimental situation led to a good agreement of the simulated cumulative distribution functions of the search time with the experimentally computed equivalent.

The experimentally measured mean search time of CadC for its DNA binding site was found to be independent of the position of the binding site along the chromosome. Adding a second binding site however led to a reduction of the search time by a factor of two. Both observations are in agreement with our numerical simulations, which shows that the reduction of search time is a result of the existence of two independent and equally accessible binding sites, where distant parts of the polymer become uncorrelated in their motion due to the confinement in the cell. While in the case of CadC there is only a single specific binding site and the number of proteins is highly controlled to 1–3 molecules [184], i.e. a single dimer, the inverse proportionality of the mean search time to the number of independent binding sites and proteins provides another way to speed up the process.

Varying the diffusion of the protein and the polymer in our simulations showed that in the experimentally relevant regime for membrane-integrated proteins the search time is inversely proportional to the diffusion coefficient of the protein, but relatively insensitive to the polymer diffusion speed. The search process therefore seems to be limited by the mobility of the protein, while the polymer has to be just mobile enough for the binding site to randomly come close to the membrane. Having a very fast diffusing polymer like a

small mRNA molecule however could lead to a different search mechanism as the polymer movement could play a more central role.

While the experiments on the target search of CadC in *E. coli* are in excellent agreement with our simulations, it could still be of interest to extend the fluorescence microscopy protocol to simultaneously label CadC and its cognate binding site on the DNA. Especially if it would be conceivable to obtain time-lapse videos of the search process, the relative motion of CadC and its binding site could be observed directly. Time-lapse videos could also answer the question how far apart CadC and its binding site are at the onset of activating conditions and where the CadC foci move at cell division. To test the results from our numerical simulations concerning the dependence on polymer density *in vivo* is not easily conceivable, as placing the *cadBA* promoter on a plasmid of varying size has little influence on the total polymer density in the cell. What could be tested easily however is the effect of more than two binding sites on the chromosome, as well as varying the number of CadC molecules in the membrane.

As our simulations are in very good agreement with the fluorescence microscopy experiments of the *E. coli* CadC target search dynamics, they could provide a way to better understand the kinetics of similar search processes. A well suited example is CadC in *Vibrio campbellii*, which shows significant homology to CadC in *E. coli*. As the copy number of CadC molecules per cell in *V. campbellii* is about 11 under normal conditions and increases by a factor of five under acid stress conditions [203], this provides an interesting target search problem to study. Since our simulations have shown that the target search becomes independent of polymer length at high polymer densities, it could also be conceivable to extend the polymer model to consider the viscoelasticity of the cytoplasm like in the Brownian dynamics simulations with a memory kernel established in a recent study [204]. Moreover, the simulated model could be extended to include cell growth and chromosome segregation. An example of a target search process where this plays a central role is the lysogeny pathway of λ bacteriophage infection. After injection of its DNA into the cell, integration of the viral DNA by site-specific recombination requires the encounter of the two attachment sites *attB* and *attP* in the bacterial and the phage genome respectively. Measurements of integration frequencies in dependence on the location of *attB* in *E. coli* have revealed an interesting search strategy of the bacteriophage [205]. The λ DNA remains close to its entry point at the pole or at midcell of the host and uses the directed motion of the chromosome during segregation to locate its specific insertion sequence, thus providing a prime example of coevolution of λ with its host.

Besides CadC in *E. coli* members of the ToxR receptor family also include its homologue in *V. cholerae* [25]. The epidemic diarrheal disease cholera is caused by *V. cholerae*, which enters orally and passes the low-pH environment of the stomach to colonise in the intestine. Since reducing the ability to survive acid exposure can lead to decreased virulence, a thorough understanding of the functioning of CadC can be of interest to drug design, all the more as the major virulence gene regulator ToxR is also a one-component receptor. The ability of bacteria to sense an extracellular signal and transduce it into a cellular response also harbours the potential of reprogramming bacteria as biosensors [206]. Programming bacterial biosensors to monitor the concentration of toxins or to detect an infection and produce a toxin to kill the pathogen requires the detailed understanding of existing signalling pathways like the one-component receptors studied in this dissertation.

Taken together, our findings predict that for a search process in a low polymer density environment, like mRNA molecules searching for nuclear pore complexes, transient non-specific binding could be relevant to speed up the process. Moreover, both the large number of NPCs and the small size of mRNA molecules lead to a considerable speed-up. For membrane-integrated transcription factors in bacteria however, where polymer density is high, a decorrelation of polymer subchains due to the confinement facilitates the search process, rendering elaborate search strategies unnecessary even for low copy number receptors like CadC in *E. coli*, that have to locate a single cognate binding site on the chromosome.

Bibliography V

[1] E. Machery. Why i stopped worrying about the definition of life... and why you should as well. *Synthese*, 185(1):145–164, 2012.

[2] E. Bianconi, A. Piovesan, F. Facchin, A. Beraudi, R. Casadei, F. Frabetti, L. Vitale, M. C. Pelleri, S. Tassani, F. Piva, et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.*, 40(6):463–471, 2013.

[3] R. F. Bruinsma. Physics of protein-dna interaction. In *Physics of bio-molecules and cells. Physique des biomolécules et des cellules*, pages 1–68. Springer, 2002.

[4] K. Roberts, B. Alberts, A. Johnson, P. Walter, and T. Hunt. Molecular biology of the cell. *New York: Garland Science*, 2002.

[5] S. C. Verma, Z. Qian, and S. L. Adhya. Architecture of the escherichia coli nucleoid. *PLos Genet.*, 15(12):e1008456, 2019.

[6] B. Youngren, H. J. Nielsen, S. Jun, and S. Austin. The multifork escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes. Dev.*, 28(1):71–84, 2014.

[7] Z. D. Blount. The natural history of model organisms: The unexhausted potential of e. coli. *Elife*, 4:e05826, 2015.

[8] J. B. Kaper, J. P. Nataro, and H. L. Mobley. Pathogenic escherichia coli. *Nat. Rev. Microbiol.*, 2(2):123–140, 2004.

[9] X. Wang, P. M. Llopis, and D. Z. Rudner. Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.*, 14(3):191–203, 2013.

[10] A. Badrinarayanan, T. B. Le, and M. T. Laub. Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.*, 31:171–199, 2015.

[11] R. T. Dame, O. J. Kalmykowa, and D. C. Grainger. Chromosomal macrodomains and associated proteins: implications for dna organization and replication in gram negative bacteria. *PLos Genet.*, 7(6):e1002123, 2011.

[12] X. Wang and D. Z. Rudner. Spatial organization of bacterial chromosomes. *Curr. Opin. Microbiol.*, 22:66–72, 2014.

[13] M. Lewis. The lac repressor. *C.R. Biol.*, 328(6):521–548, 2005.

[14] S. Busby and R. H. Ebright. Transcription activation by catabolite activator protein (cap). *J. Mol. Biol.*, 293(2):199–213, 1999.

[15] C. M. Guldberg and P. Waage. *Etudes sur les affinités chimiques.* Imprimerie de Brøgger & Christie, 1867.

[16] M. Von Smoluchowski. Mathematical theory of the kinetics of the coagulation of colloidal solutions. *Z. Phys. Chem*, 92(129):129–68, 1917.

[17] J. M. Schurr. The role of diffusion in bimolecular solution kinetics. *Biophys. J.*, 10(8):700–716, 1970.

[18] L. E. Ulrich, E. V. Koonin, and I. B. Zhulin. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.*, 13(2):52–56, 2005.

[19] A. M. Stock, V. L. Robinson, and P. N. Goudreau. Two-component signal transduction. *Annu. Rev. Biochem.*, 69(1):183–215, 2000.

[20] V. L. Miller, R. K. Taylor, and J. J. Mekalanos. Cholera toxin transcriptional activator toxr is a transmembrane dna binding protein. *Cell*, 48(2):271–279, 1987.

[21] N. C. Shaner, G. G. Lambert, A. Chammas, Y. Ni, P. J. Cranfill, M. A. Baird, B. R. Sell, J. R. Allen, R. N. Day, M. Israelsson, et al. A bright monomeric green fluorescent protein derived from branchiostoma lanceolatum. *Nat. Methods*, 10(5):407, 2013.

[22] Y. Yang and R. R. Isberg. Transcriptional regulation of the yersinia pseudotuberculosis ph 6 antigen adhesin by two envelope-associated components. *Mol. Microbiol.*, 24(3):499–510, 1997.

[23] S. Stelzer, S. Egan, M. R. Larsen, D. H. Bartlett, and S. Kjelleberg. Unravelling the role of the toxr-like transcriptional regulator wmpr in the marine antifouling bacterium pseudoalteromonas tunicata. *Microbiol.*, 152(5):1385–1394, 2006.

[24] S. Buchner, A. Schlundt, J. Lassak, M. Sattler, and K. Jung. Structural and functional analysis of the signal-transducing linker in the ph-responsive one-component system cadc of escherichia coli. *J. Mol. Biol.*, 427(15):2548–2561, 2015.

[25] D. S. Merrell and A. Camilli. Regulation of vibrio cholerae genes required for acid tolerance by a member of the "toxr-like" family of transcriptional regulators. *J. Bacteriol.*, 182(19):5342–5350, 2000.

[26] J. H. Cho, S. J. Ha, L. R. Kao, T. L. Megraw, and C.-B. Chae. A novel dna-binding protein bound to the mitochondrial inner membrane restores the null mutation of mitochondrial histone abf2p in saccharomyces cerevisiae. *Mol. Cell. Biol.*, 18(10):5712–5723, 1998.

[27] R. K. Thauer. Citric-acid cycle, 50 years on: Modifications and an alternative pathway in anaerobic bacteria. *Eur. J. Biochem.*, 176(3):497–508, 1988.

[28] V. Iacobazzi and V. Infantino. Citrate–new functions for an old metabolite. *Biol. Chem.*, 395(4):387–399, 2014.

[29] M. Beck, F. Förster, M. Ecke, J. M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister, and O. Medalia. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science*, 306(5700):1387–1390, 2004.

[30] L. I. Davis. The nuclear pore complex. *Annu. Rev. Biochem.*, 64(1):865–896, 1995.

[31] C. H. Freudenreich and X. A. Su. Relocalization of dna lesions to the nuclear pore complex. *FEMS Yeast Res.*, 16(8):fow095, 2016.

[32] S. Nagai, K. Dubrana, M. Tsai-Pflugfelder, M. B. Davidson, T. M. Roberts, G. W. Brown, E. Varela, F. Hediger, S. M. Gasser, and N. J. Krogan. Functional targeting of dna damage to a nuclear pore-associated sumo-dependent ubiquitin ligase. *Science*, 322(5901):597–602, 2008.

[33] G. Blobel. Gene gating: a hypothesis. *Proc. Natl. Acad. Sci.*, 82(24):8527–8529, 1985.

[34] L. T. Burns and S. R. Wente. From hypothesis to mechanism: uncovering nuclear pore complex links to gene expression. *Mol. Cell. Biol.*, 34(12):2114–2120, 2014.

[35] V. Sood and J. H. Brickner. Nuclear pore interactions with the genome. *Curr. Opin. Genet. Dev.*, 25:43–49, 2014.

[36] A. D. Riggs, S. Bourgeois, and M. Cohn. The lac represser-operator interaction: Iii. kinetic studies. *J. Mol. Biol.*, 53(3):401–417, 1970.

[37] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj. How a protein searches for its site on dna: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.*, 42(43):434013, 2009.

[38] G. Adam and M. Delbrück. Reduction of dimensionality in biological diffusion processes. *Struct. Chem. Mol. Biol.*, 198:198–215, 1968.

[39] A. Revzin. *The biology of nonspecific DNA protein interactions.* CRC Press, 1990.

[40] P. H. Richter and M. Eigen. Diffusion controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes. *Biophys. Chem.*, 2(3):255–263, 1974.

[41] O. G. Berg, R. B. Winter, and P. H. Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry (Mosc.)*, 20(24):6929–6948, 1981.

[42] R. B. Winter and P. H. Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. the escherichia coli lac repressor-operator interaction: equilibrium measurements. *Biochemistry (Mosc.)*, 20(24):6948–6960, 1981.

[43] R. B. Winter, O. G. Berg, and P. H. Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. the escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry (Mosc.)*, 20(24):6961–6977, 1981.

[44] M. Slutsky and L. A. Mirny. Kinetics of protein-dna interaction: facilitated target location in sequence-dependent potential. *Biophys. J.*, 87(6):4021–4035, 2004.

[45] T. Hu, A. Y. Grosberg, and B. Shklovskii. How proteins search for their specific sites on dna: the role of dna conformation. *Biophys. J.*, 90(8):2731–2744, 2006.

[46] D. M. Gowers, G. G. Wilson, and S. E. Halford. Measurement of the contributions of 1d and 3d pathways to the translocation of a protein along dna. *Proc. Natl. Acad. Sci.*, 102(44):15883–15888, 2005.

[47] I. Bonnet, A. Biebricher, P.-L. Porte, C. Loverdo, O. Bénichou, R. Voituriez, C. Escude, W. Wende, A. Pingoud, and P. Desbiolles. Sliding and jumping of single ecorv restriction enzymes on non-cognate dna. *Nucleic Acids Res.*, 36(12):4118–4127, 2008.

[48] B. van den Broek, M. A. Lomholt, S.-M. Kalisch, R. Metzler, and G. J. Wuite. How dna coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci.*, 105(41):15738–15742, 2008.

[49] J. Gorman, A. J. Plys, M.-L. Visnapuu, E. Alani, and E. C. Greene. Visualizing one-dimensional diffusion of eukaryotic dna repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, 17(8):932–938, 2010.

[50] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf. The lac repressor displays facilitated diffusion in living cells. *Science*, 336(6088):1595–1598, 2012.

[51] J. Elf, G.-W. Li, and X. S. Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, 2007.

[52] A. Veksler and A. B. Kolomeisky. Speed-selectivity paradox in the protein search for targets on dna: is it real or not? *J. Phys. Chem. B*, 117(42):12695–12701, 2013.

[53] M. Bauer and R. Metzler. Generalized facilitated diffusion model for dna-binding proteins with search and recognition states. *Biophys. J.*, 102(10):2321–2330, 2012.

[54] L. Cuculis, Z. Abil, H. Zhao, and C. M. Schroeder. Direct observation of tale protein dynamics reveals a two-state search mechanism. *Nat. Commun.*, 6(1):1–11, 2015.

[55] E. G. Marklund, A. Mahmutovic, O. G. Berg, P. Hammar, D. van der Spoel, D. Fange, and J. Elf. Transcription-factor binding and sliding on dna studied using micro-and macroscopic models. *Proc. Natl. Acad. Sci.*, 110(49):19796–19801, 2013.

[56] C. A. Brackley, M. E. Cates, and D. Marenduzzo. Facilitated diffusion on mobile dna: configurational traps and sequence heterogeneity. *Phys. Rev. Lett.*, 109(16):168103, 2012.

[57] T. Hu, A. Y. Grosberg, and B. Shklovskii. How proteins search for their specific sites on dna: the role of dna conformation. *Biophys. J.*, 90(8):2731–2744, 2006.

[58] T. Schötz, R. A. Neher, and U. Gerland. Target search on a dynamic dna molecule. *Phys. Rev. E*, 84(5):051911, 2011.

[59] G.-W. Li, O. G. Berg, and J. Elf. Effects of macromolecular crowding and dna looping on gene regulation kinetics. *Nat. Phys.*, 5(4):294, 2009.

[60] C. Brackley, M. Cates, and D. Marenduzzo. Intracellular facilitated diffusion: searchers, crowders, and blockers. *Phys. Rev. Lett.*, 111(10):108101, 2013.

[61] H. v. Helmholtz. *Theorie der Luftschwingungen in Röhren mit offenen Enden.* De Gruyter, 2020.

[62] J. W. Strutt and J. W. S. B. Rayleigh. *The theory of sound*, volume 1. Macmillan, 1877.

[63] A. Singer, Z. Schuss, D. Holcman, and R. S. Eisenberg. Narrow escape, part i. *J. Stat. Phys.*, 122(3):437–463, 2006.

[64] A. Singer, Z. Schuss, and D. Holcman. Narrow escape, part iii: Non-smooth domains and riemann surfaces. *J. Stat. Phys.*, 122(3):491–509, 2006.

[65] A. Eichinger, I. Haneburger, C. Koller, K. Jung, and A. Skerra. Crystal structure of the sensory domain of escherichia coli cadc, a member of the toxr-like protein family. *Protein Sci.*, 20(4):656–669, 2011.

[66] A. Singer, Z. Schuss, and D. Holcman. Narrow escape, part iii: Non-smooth domains and riemann surfaces. *J. Stat. Phys.*, 122(3):491–509, 2006.

[67] A. Singer, Z. Schuss, and D. Holcman. Narrow escape, part ii: The circular disk. *J. Stat. Phys.*, 122(3):465–489, 2006.

[68] S. Pillay, M. J. Ward, A. Peirce, and T. Kolokolnikov. An asymptotic analysis of the mean first passage time for narrow escape problems: Part i: Two-dimensional domains. *Multiscale Model. Sim.*, 8(3):803–835, 2010.

[69] A. F. Cheviakov, M. J. Ward, and R. Straube. An asymptotic analysis of the mean first passage time for narrow escape problems: Part ii: The sphere. *Multiscale Model. Sim.*, 8(3):836–870, 2010.

[70] C. L. Woldringh. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol. Microbiol.*, 45(1):17–29, 2002.

[71] M. Roggiani and M. Goulian. Chromosome-membrane interactions in bacteria. *Annu. Rev. Genet.*, 49:115–129, 2015.

[72] M. A. P. Rodriguez and X. Guo. Biomacromolecular localization in bacterial cells by the diffusion and capture mechanism. *Ann. Microbiol.*, 63(3):825–832, 2013.

[73] V. Norris. Hypothesis: chromosome separation in escherichia coli involves autocatalytic gene expression, transertion and membrane-domain formation. *Mol. Microbiol.*, 16(6):1051–1057, 1995.

[74] K. Matsumoto, H. Hara, I. Fishov, E. Mileykovskaya, and V. Norris. The membrane: transertion as an organizing principle in membrane heterogeneity. *Front. Microbiol.*, 6:572, 2015.

[75] E. A. Libby, M. Roggiani, and M. Goulian. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proc. Natl. Acad. Sci.*, 109(19):7445–7450, 2012.

[76] S. Bakshi, H. Choi, J. Mondal, and J. C. Weisshaar. Time-dependent effects of transcription-and translation-halting drugs on the spatial distributions of the e scherichia coli chromosome and ribosomes. *Mol. Microbiol.*, 94(4):871–887, 2014.

[77] S. Urban and M. Freeman. Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr. Opin. Genet. Dev.*, 12(5):512–518, 2002.

[78] Y. H. Lee, J. H. Kim, I. S. Bang, and Y. K. Park. The membrane-bound transcriptional regulator cadc is activated by proteolytic cleavage in response to acid stress. *J. Bacteriol.*, 190(14):5120–5126, 2008.

[79] M. Thanbichler and L. Shapiro. Getting organized—how bacterial cells move proteins and dna. *Nat. Rev. Microbiol.*, 6(1):28–40, 2008.

[80] D. Z. Rudner, Q. Pan, and R. M. Losick. Evidence that subcellular localization of a bacterial membrane protein is achieved by diffusion and capture. *Proc. Natl. Acad. Sci.*, 99(13):8701–8706, 2002.

[81] S. Brameyer, T. C. Rösch, J. El Andari, E. Hoyer, J. Schwarz, P. L. Graumann, and K. Jung. Dna-binding directs the localization of a membrane-integrated receptor of the toxr family. *Commun. Biol.*, 2(1):4, 2019.

[82] G. Fritz, C. Koller, K. Burdack, L. Tetsch, I. Haneburger, K. Jung, and U. Gerland. Induction kinetics of a conditional ph stress response system in escherichia coli. *J. Mol. Biol.*, 393(2):272–286, 2009.

[83] I. Haneburger, G. Fritz, N. Jurkschat, L. Tetsch, A. Eichinger, A. Skerra, U. Gerland, and K. Jung. Deactivation of the e. coli ph stress sensor cadc by cadaverine. *J. Mol. Biol.*, 424(1-2):15–27, 2012.

[84] S. M. Hoseinpoor, N. Nikoofard, and B.-Y. Ha. Characteristic time for the end monomers of a spherically confined polymer to find a nano-pore. *J. Chem. Phys.*, 154(11):114901, 2021.

[85] A. Amitai, C. Amoruso, A. Ziskind, and D. Holcman. Encounter dynamics of a small target by a polymer diffusing in a confined domain. *J. Chem. Phys.*, 137(24):244906, 2012.

[86] J. M. Polson and Z. R. McLure. Free-energy cost of localizing a single monomer of a confined polymer. *Phys. Rev. E*, 99(6):062503, 2019.

[87] S. Ramadurai, A. Holt, V. Krasnikov, G. van den Bogaart, J. A. Killian, and B. Poolman. Lateral diffusion of membrane proteins. *J. Am. Chem. Soc.*, 131(35):12650–12656, 2009.

[88] T. Kawakatsu. *Statistical physics of polymers: an introduction*. Springer Science & Business Media, 2013.

[89] P.-G. De Gennes. *Scaling concepts in polymer physics*. Cornell university press, 1979.

[90] M. Rubinstein, R. H. Colby, et al. *Polymer physics*, volume 23. Oxford university press New York, 2003.

[91] G. R. Strobl and G. R. Strobl. *The physics of polymers*, volume 2. Springer, 1997.

[92] J. Kierfeld, K. Baczynski, P. Gutjahr, and R. Lipowsky. Semiflexible polymers and filaments: From variational problems to fluctuations. In *AIP Conference Proceedings*, volume 1002, pages 151–185. American Institute of Physics, 2008.

[93] W. Kuhn. über die gestalt fadenförmiger moleküle in lösungen. *Kolloid-Zeitschrift*, 68(1):2–15, 1934.

[94] M. Doi and S. F. Edwards. *The theory of polymer dynamics*, volume 73. oxford university press, 1988.

[95] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith. Single-molecule studies of dna mechanics. *Curr. Opin. Struct. Biol.*, 10(3):279–285, 2000.

[96] O. Kratky and G. Porod. Röntgenuntersuchung gelöster fadenmoleküle. *Recl. Trav. Chim. Pays-Bas*, 68(12):1106–1122, 1949.

[97] G. R. Strobl. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior*. Springer Science & Business Media, 2007.

[98] R. Brown. Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philos. Mag.*, 4(21):161–173, 1828.

[99] A. Einstein et al. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Ann. Phys.*, 17(549-560):208, 1905.

[100] M. Smoluchowski. The kinetic theory of brownian molecular motion and suspensions. *Ann. Phys*, 21:756–780, 1906.

[101] A. Fick. V. on liquid diffusion. *Lond. Edinb. Dubl. Phil. Mag.*, 10(63):30–39, 1855.

[102] M. Doi and S. F. Edwards. *The theory of polymer dynamics*, volume 73. oxford university press, 1988.

[103] L. Onsager. Reciprocal relations in irreversible processes. i. *Physs Rev.*, 37(4):405, 1931.

[104] P. Langevin. Sur la théorie du mouvement brownien. *Compt. Rendus*, 146:530–533, 1908.

[105] P. E. Rouse Jr. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21(7):1272–1280, 1953.

[106] D. Michieletto. Notes on polymers. 2016.

[107] J. T. Padding. Theory of polymer dynamics. *Advanced Courses in Macroscopic Physical Chemistry.(Han-sur-lesse winterschool 2005) https://www. scribd. com/document/181367522/THEORYOF-POLYMER-DYNAMICS-Paddings*, 21, 2005.

[108] B. H. Zimm. Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.*, 24(2):269–278, 1956.

[109] S. R. Quake. The zimm model applied to extended single polymers. *J. Chem. Phys.*, 101(5):4307–4311, 1994.

[110] N. Metropolis and S. Ulam. The monte carlo method. *J. Am. Stat. Assoc.*, 44(247):335–341, 1949.

[111] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[112] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27(5):1208–1209, 1957.

[113] G. Harp, J. Miller, and B. J. Berne. Attainment of statistical equilibrium in excited nuclei. *Phys. Rev.*, 165(4):1166, 1968.

[114] J. Barker and R. Watts. Structure of water; a monte carlo calculation. *Chem. Phys. Lett.*, 3(3):144–145, 1969.

[115] A. Rahman and F. H. Stillinger. Molecular dynamics study of liquid water. *J. Chem. Phys.*, 55(7):3336–3359, 1971.

[116] K. Binder. *Monte Carlo and molecular dynamics simulations in polymer science*. Oxford University Press, 1995.

[117] L. Monticelli and E. Salonen. *Biomolecular simulations: methods and protocols*, volume 924. Springer, 2013.

[118] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[119] S. Marrink, D. Tieleman, and A. Mark. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *J. Phys. Chem. B*, 104(51):12165–12173, 2000.

[120] D. E. Shaw, R. O. Dror, J. K. Salmon, J. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, et al. Millisecond-scale molecular dynamics simulations on anton. In *Proceedings of the conference on high performance computing networking, storage and analysis*, pages 1–11, 2009.

[121] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[122] D. L. Ermak and Y. Yeh. Equilibrium electrostatic effects on the behavior of polyions in solution: Polyion-mobile ion interaction. *Chem. Phys. Lett.*, 24(2):243–248, 1974.

[123] J. Weiner and R. Forman. Rate theory for solids. iv. classical brownian-motion model. *Phys. Rev. B*, 10(2):315, 1974.

[124] P. Turq, F. Lantelme, and H. L. Friedman. Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.*, 66(7):3039–3044, 1977.

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

[126] M. Pear and J. Weiner. Brownian dynamics study of a polymer chain of linked rigid bodies. *J. Chem. Phys.*, 71(1):212–224, 1979.

[127] S. G. Hatzikiriakos and D. Vlassopoulos. Brownian dynamics simulations of shear-thickening in dilute polymer solutions. *Rheol. Acta*, 35(3):274–287, 1996.

[128] J. D. Dwyer and V. A. Bloomfield. Brownian dynamics simulations of probe and self-diffusion in concentrated protein and dna solutions. *Biophys. J.*, 65(5):1810–1816, 1993.

[129] H. Merlitz, K. Rippe, K. V. Klenin, and J. Langowski. Looping dynamics of linear dna molecules and the effect of dna curvature: a study by brownian dynamics simulation. *Biophys. J.*, 74(2):773–779, 1998.

[130] W. Young and E. Elcock. Monte carlo studies of vacancy migration in binary ordered alloys: I. *Proceedings of the Physical Society (1958-1967)*, 89(3):735, 1966.

[131] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for monte carlo simulation of ising spin systems. *J. Comput. Phys.*, 17(1):10–18, 1975.

[132] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.

[133] B. Sauerwine and M. Widom. Kinetic monte carlo method applied to nucleic acid hairpin folding. *Phys. Rev. E*, 84(6):061912, 2011.

[134] M. Kotrla. Numerical simulations in the theory of crystal growth. *Comput. Phys. Commun.*, 97(1-2):82–100, 1996.

[135] A. Rey and J. Skolnick. Comparison of lattice monte carlo dynamics and brownian dynamics folding pathways of α-helical hairpins. *Chem. Phys.*, 158(2-3):199–219, 1991.

[136] A. Kolinski, M. Milik, and J. Skolnick. Static and dynamic properties of a new lattice model of polypeptide chains. *J. Chem. Phys.*, 94(5):3978–3985, 1991.

[137] A. Baumgärtner, K. Kremer, and K. Binder. Dynamics of entangled flexible polymers. monte carlo simulations and their interpretation. In *Faraday Symposia of the Chemical Society*, volume 18, pages 37–47. Royal Society of Chemistry, 1983.

[138] M. Levitt. Computer simulation of dna double-helix dynamics. In *Cold Spring Harbor symposia on quantitative biology*, volume 47, pages 251–262. Cold Spring Harbor Laboratory Press, 1983.

[139] M. W. Konrad and J. I. Bolonick. Molecular dynamics simulation of dna stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *J. Am. Chem. Soc.*, 118(45):10989–10994, 1996.

[140] J. Mitchell, C. Laughton, and S. A. Harris. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled dna. *Nucleic Acids Res.*, 39(9):3928–3938, 2011.

[141] G. Chirico and J. Langowski. Kinetics of dna supercoiling studied by brownian dynamics simulation. *Biopolymers*, 34(3):415–433, 1994.

[142] K. Ott, L. Martini, J. Lipfert, and U. Gerland. Dynamics of the buckling transition in double-stranded dna and rna. *Biophys. J.*, 118(7):1690–1701, 2020.

[143] C. A. Miermans and C. P. Broedersz. A lattice kinetic monte-carlo method for simulating chromosomal dynamics and other (non-) equilibrium bio-assemblies. *Soft Matter*, 16(2):544–556, 2020.

[144] R. J. Menssen, G. J. Kimmel, and A. Tokmakoff. Investigation into the mechanism and dynamics of dna association and dissociation utilizing kinetic monte carlo simulations. *J. Chem. Phys.*, 154(4):045101, 2021.

[145] C.-C. Hsieh, S. Jain, and R. G. Larson. Brownian dynamics simulations with stiff finitely extensible nonlinear elastic-fraenkel springs as approximations to rods in bead-rod models. *J. Chem. Phys.*, 124(4):044911, 2006.

[146] S. Geggier, A. Kotlyar, and A. Vologodskii. Temperature dependence of dna persistence length. *Nucleic Acids Res.*, 39(4):1419–1426, 2011.

[147] P. T. Underhill and P. S. Doyle. On the coarse-graining of polymers into bead-spring chains. *J. Non-Newtonian Fluid Mech.*, 122(1-3):3–31, 2004.

[148] L. Euler. *Institutiones calculi integralis*, volume 4. Academia Imperialis Scientiarum, 1794.

[149] A. Ladd. Numerical methods for molecular and continuum dynamics. *Lectures at the 3rd Warsaw School of Statistical Physics (Kazimierz, Poland)[http://www. che. ufl. edu/ladd/publications/kmz-09. pdf]*, 2009.

[150] W. Möbius, E. Frey, and U. Gerland. Spontaneous unknotting of a polymer confined in a nanochannel. *Nano Lett.*, 8(12):4518–4522, 2008.

[151] A. Prados, J. Brey, and B. Sánchez-Rey. A dynamical monte carlo algorithm for master equations with time-dependent transition rates. *J. Stat. Phys.*, 89(3):709–734, 1997.

[152] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.

[153] A. Chatterjee and D. G. Vlachos. An overview of spatial microscopic and accelerated kinetic monte carlo methods. *J. Comput. Aided Mol. Des.*, 14(2):253–308, 2007.

[154] A. F. Voter. Introduction to the kinetic monte carlo method. In *Radiation effects in solids*, pages 1–23. Springer, 2007.

[155] A. Chatterjee and D. G. Vlachos. An overview of spatial microscopic and accelerated kinetic monte carlo methods. *J. Comput. Aided Mol. Des.*, 14(2):253–308, 2007.

[156] G. Gilmer and P. Bennema. Simulation of crystal growth with surface diffusion. *J. Appl. Phys.*, 43(4):1347–1360, 1972.

[157] P. H. Verdier and W. Stockmayer. Monte carlo calculations on the dynamics of polymers in dilute solution. *J. Chem. Phys.*, 36(1):227–235, 1962.

[158] A. Baumgartner. Simulation of polymer motion. *Annu. Rev. Phys. Chem.*, 35(1):419–435, 1984.

[159] H. Hilhorst and J. Deutch. Analysis of monte carlo results on the kinetics of lattice polymer chains with excluded volume. *J. Chem. Phys.*, 63(12):5153–5161, 1975.

[160] F. T. Wall and F. Mandel. Macromolecular dimensions obtained by an efficient monte carlo method without sample attrition. *J. Chem. Phys.*, 63(11):4592–4595, 1975.

[161] M. Lal. 'monte carlo'computer simulation of chain molecules. i. *Mol. Phys.*, 17(1):57–64, 1969.

[162] I. Carmesin and K. Kremer. The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. *Macromolecules*, 21(9):2819–2823, 1988.

[163] H.-P. Wittmann and K. Kremer. Vectorized version of the bond fluctuation method for lattice polymers. *Comput. Phys. Commun.*, 61(3):309–330, 1990.

[164] O. Hassager. Kinetic theory and rheology of bead-rod models for macromolecular solutions. i. equilibrium and steady flow properties. *J. Chem. Phys.*, 60(5):2111–2124, 1974.

[165] S. Shanbhag and R. G. Larson. Chain retraction potential in a fixed entanglement network. *Phys. Rev. Lett.*, 94(7):076001, 2005.

[166] J. Wittmer, A. Cavallo, T. Kreer, J. Baschnagel, and A. Johner. A finite excluded volume bond-fluctuation model: Static properties of dense polymer melts revisited. *J. Chem. Phys.*, 131(6):064901, 2009.

[167] G. Tiana, L. Sutto, and R. A. Broglia. Use of the metropolis algorithm to simulate the dynamics of protein chains. *Physica A*, 380:241–249, 2007.

[168] L. Martini and U. Gerland. Search dynamics of a confined polymer. In preparation.

[169] O. Benichou, T. Guérin, and R. Voituriez. Mean first-passage times in confined media: from markovian to non-markovian processes. *J. Phys. A: Math. Theor.*, 48(16):163001, 2015.

[170] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer, and R. Voituriez. Geometry-controlled kinetics. *Nat. Chem.*, 2(6):472–477, 2010.

[171] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, January 1968.

[172] T. Guérin, O. Bénichou, and R. Voituriez. Reactive conformations and non-markovian cyclization kinetics of a rouse polymer. *J. Chem. Phys.*, 138(9):094908, 2013.

[173] M. F. Maghrebi, Y. Kantor, and M. Kardar. Entropic force of polymers on a cone tip. *Europhys. Lett.*, 96(6):66002, 2011.

[174] I. V. Grigoriev, Y. A. Makhnovskii, A. M. Berezhkovskii, and V. Y. Zitserman. Kinetics of escape through a small hole. *J. Chem. Phys.*, 116(22):9574–9577, 2002.

[175] G. G. Maul and L. Deaven. Quantitative determination of nuclear pore complexes in cycling cells with differing dna content. *J. Cell Biol.*, 73(3):748–760, 1977.

[176] L. M. Garcia-Segura, M. Lafarga, M. Berciano, P. Hernandez, and M. Andres. Distribution of nuclear pores and chromatin organization in neurons and glial cells of the rat cerebellar cortex. *J. Comp. Neurol.*, 290(3):440–450, 1989.

[177] L. Martini, S. Brameyer, E. Hoyer, K. Jung, and U. Gerland. Dynamics of chromosomal target search by a membrane-integrated one-component receptor. *PLoS Comput. Biol.*, 17(2):e1008680, 2021.

[178] L. Hall-Stoodley, J. W. Costerton, and P. Stoodley. Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.*, 2(2):95, 2004.

[179] A. Schlundt, S. Buchner, R. Janowski, T. Heydenreich, R. Heermann, J. Lassak, A. Geerlof, R. Stehle, D. Niessing, K. Jung, and M. Sattler. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Sci. Rep.*, 7(1):1051, 2017.

[180] I. Haneburger, A. Eichinger, A. Skerra, and K. Jung. New insights into the signaling mechanism of the ph-responsive, membrane-integrated transcriptional activator cadc of escherichia coli. *J. Biol. Chem.*, 286(12):10681–10689, 2011.

[181] L. Tetsch, C. Koller, I. Haneburger, and K. Jung. The membrane-integrated transcriptional activator cadc of escherichia coli senses lysine indirectly via the interaction with the lysine permease lysp. *Mol. Microbiol.*, 67(3):570–583, 2008.

[182] M. Rauschmeier, V. Schüppel, L. Tetsch, and K. Jung. New insights into the interplay between the lysine transporter lysp and the ph sensor cadc in escherichia coli. *J. Mol. Biol.*, 426(1):215–229, 2014.

[183] E. Lindner and S. H. White. Topology, dimerization, and stability of the single-span membrane protein cadc. *J. Mol. Biol.*, 426(16):2942–2957, 2014.

[184] S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson, and K. Jung. Translation elongation factor ef-p alleviates ribosome stalling at polyproline stretches. *Science*, 339(6115):82–85, 2013.

[185] B. Görke, J. Reinhardt, and B. Rak. Activity of lac repressor anchored to the escherichia coli inner membrane. *Nucleic Acids Res.*, 33(8):2504–2511, 2005.

[186] S. Redner. *A guide to first-passage processes.* Cambridge University Press, 2001.

[187] J. Zhou and K. E. Rudd. Ecogene 3.0. *Nucleic Acids Res.*, 41(D1):D613–D624, 2012.

[188] A. Paintdakhi, B. Parry, M. Campos, I. Irnov, J. Elf, I. Surovtsev, and C. Jacobs-Wagner. Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.*, 99(4):767–777, 2016.

[189] R. Prats and M. De Pedro. Normal growth and division of escherichia coli with a reduced amount of murein. *J. Bacteriol.*, 171(7):3740–3745, 1989.

[190] H. P. Erickson. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online*, 11(1):32, 2009.

[191] N. Baranova, P. Radler, V. M. Hernández-Rocamora, C. Alfonso, M. López-Pelegrín, G. Rivas, W. Vollmer, and M. Loose. Diffusion and capture permits dynamic coupling between treadmilling ftsz filaments and cell division proteins. *Nat. Microbiol.*, 5(3):407–417, 2020.

[192] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, et al. The complete genome sequence of escherichia coli k-12. *Science*, 277(5331):1453–1462, 1997.

[193] M. Kumar, M. S. Mommer, and V. Sourjik. Mobility of cytoplasmic, membrane, and dna-binding proteins in escherichia coli. *Biophys. J.*, 98(4):552–559, 2010.

[194] F. Oswald, A. Varadarajan, H. Lill, E. J. Peterman, and Y. J. Bollen. Mreb-dependent organization of the e. coli cytoplasmic membrane controls membrane protein diffusion. *Biophys. J.*, 110(5):1139–1149, 2016.

[195] F. Oswald, E. L. Bank, Y. J. Bollen, and E. J. Peterman. Imaging and quantification of trans-membrane protein diffusion in living bacteria. *Phys. Chem. Chem. Phys.*, 16(25):12625–12634, 2014.

[196] S. C. Weber, A. J. Spakowitz, and J. A. Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.*, 104(23):238102, 2010.

[197] O. Espeli, R. Mercier, and F. Boccard. Dna dynamics vary according to macrodomain topography in the e. coli chromosome. *Mol. Microbiol.*, 68(6):1418–1427, 2008.

[198] A. Javer, Z. Long, E. Nugent, M. Grisi, K. Siriwatwetchakul, K. D. Dorfman, P. Cicuta, and M. C. Lagomarsino. Short-time movement of e. coli chromosomal loci depends on coordinate and subcellular localization. *Nat. Commun.*, 4:3003, 2013.

[199] M. Socol, R. Wang, D. Jost, P. Carrivain, C. Vaillant, E. Le Cam, V. Dahirel, C. Normand, K. Bystricky, J.-M. Victor, et al. Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Res.*, 2019.

[200] D. W. Erickson, S. J. Schink, V. Patsalo, J. R. Williamson, U. Gerland, and T. Hwa. A global resource allocation strategy governs growth transition kinetics of escherichia coli. *Nature*, 551(7678):119–123, 2017.

[201] Q. Chai, S. R. Webb, Z. Wang, R. E. Dutch, and Y. Wei. Study of the degradation of a multidrug transporter using a non-radioactive pulse chase method. *Anal. Bioanal. Chem.*, 408(27):7745–7751, 2016.

[202] G. Laloux and C. Jacobs-Wagner. How do bacteria localize proteins to the cell pole? *J. Cell Sci.*, 127(1):11–19, 2014.

[203] S. Brameyer, E. Hoyer, S. Bibinger, K. Burdack, J. Lassak, and K. Jung. Molecular design of a signaling system influences noise in protein abundance under acid stress in different gammaproteobacteria. *J. Bacteriol.*, 202(16):e00121–20, 2020.

[204] S. C. Weber, J. A. Theriot, and A. J. Spakowitz. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Phys. Rev. E*, 82(1):011913, 2010.

[205] A. Tal, R. Arbel-Goren, N. Costantino, J. Stavans, et al. Location of the unique integration site on an escherichia coli chromosome by bacteriophage lambda dna in vivo. *P. Natl. A. Sci.*, 111(20):7308–7312, 2014.

[206] K. Jung, F. Fabiani, E. Hoyer, and J. Lassak. Bacterial transmembrane signalling systems and their engineering for biosensing. *Open Biol.*, 8(4):180023, 2018.

[207] G. Strang. The discrete cosine transform. *SIAM review*, 41(1):135–147, 1999.

[208] T. Kreer, J. Baschnagel, M. Müller, and K. Binder. Monte carlo simulation of long chain polymer melts: Crossover from rouse to reptation dynamics. *Macromolecules*, 34(4):1105–1117, 2001.

[209] W. Epstein and B. S. Kim. Potassium transport loci in escherichia coli k-12. *J. Bacteriol.*, 108(2):639–644, 1971.

[210] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, and H. O. Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nat. Methods*, 6(5):343, 2009.

[211] V. L. Miller and J. J. Mekalanos. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in vibrio cholerae requires toxr. *J. Bacteriol.*, 170(6):2575–2583, 1988.

[212] C. Dehio and M. Meyer. Maintenance of broad-host-range incompatibility group p and group q plasmids and transposition of tn5 in bartonella henselae following conjugal plasmid transfer from escherichia coli. *J. Bacteriol.*, 179(2):538–540, 1997.

[213] C. Küper and K. Jung. Cadc-mediated activation of the cadba promoter in escherichia coli. *J. Mol. Microbiol. Biotechnol.*, 10(1):26–39, 2005.

[214] H. J. Nielsen, J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen. The escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol. Microbiol.*, 62(2):331–338, 2006.

Polymer Dynamics and Diffusion A

A1 Diffusion Equation

We solve the Langevin equation in eq. (I.11) for $U(x) = 0$, i.e. for the diffusion of a single particle in one dimension, giving

$$\zeta \frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(t)\,.$$

we integrate

$$\int_{x(0)}^{x(t)} \zeta \mathrm{d}x(t) = \int_0^t f(t)\mathrm{d}t\,,$$

and obtain

$$x(t) - x(0) = \frac{1}{\zeta}\int_0^t f(t')\mathrm{d}t'\,.$$

Computing the mean squared average yields

$$\begin{aligned}
\langle (x(t) - x(0))^2 \rangle &= \frac{1}{\zeta^2}\left\langle \int_0^t f(t')\mathrm{d}t' \int_0^t f(s)\mathrm{d}s \right\rangle \\
&= \frac{1}{\zeta^2}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}s \langle f(t')f(s)\rangle \\
&= \frac{1}{\zeta^2}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}s 2\zeta k_\mathrm{B}T\delta(t'-s) \\
&= \frac{2k_\mathrm{B}T}{\zeta}t\,.
\end{aligned}$$

A2 Rouse Mode Analysis

A2.1 Solving the continuous Rouse model

We follow the derivation of the normal modes shown in [102]. In the continuous limit of n it holds

$$\begin{aligned}
\boldsymbol{r}_n(t) - \boldsymbol{r}_{n-1}(t) &\to \frac{\partial \boldsymbol{r}_n(t)}{\partial n} \\
\boldsymbol{r}_{n+1}(t) + \boldsymbol{r}_{n-1}(t) - 2\boldsymbol{r}_n(t) &\to \frac{\partial^2 \boldsymbol{r}_n(t)}{\partial n^2}\,,
\end{aligned}$$

and the Langevin equations describing a Rouse chain become

$$\zeta \frac{\partial \boldsymbol{r}_n(t)}{\partial t} = k_\mathrm{s} \frac{\partial^2 \boldsymbol{r}_n(t)}{\partial n^2} + \boldsymbol{f}_n(t)\,, \tag{A.1}$$

with the boundary conditions

$$\left.\frac{\partial \boldsymbol{r}_n}{\partial n}\right|_{n=0} = \left.\frac{\partial \boldsymbol{r}_n}{\partial n}\right|_{n=N} = 0\,,$$

and the random force

$$\begin{aligned}\langle \boldsymbol{f}_n(t)\rangle &= 0\\ \langle f_{n\alpha}(t) f_{m\beta}(t')\rangle &= 2\zeta k_\mathrm{B} T \delta(n-m)\delta_{\alpha\beta}\delta(t-t')\,.\end{aligned}$$

For the linear transformation of $\boldsymbol{r}_n(t)$ we make the ansatz $\boldsymbol{X}_p(t) := \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{r}_n(t)$ and choose ϕ_{pn} such that the Langevin equation becomes linear and reads

$$\zeta_p \frac{\partial \boldsymbol{X}_p(t)}{\partial t} = -k_p \boldsymbol{X}_p(t) + \boldsymbol{f}_p(t)\,,$$

i.e. such that $\boldsymbol{X}_p(t)$ are the normal modes, each capable of independent motion. Inserting the ansatz into the new Langevin equation the left hand side reads

$$\begin{aligned}\zeta_p \frac{\partial \boldsymbol{X}_p(t)}{\partial t} &= \zeta_p \frac{\partial}{\partial t} \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{r}_n(t)\\ &= \zeta_p \int_0^N \mathrm{d}n \frac{\partial \boldsymbol{r}_n(t)}{\partial t} \phi_{pn}\\ &= \zeta_p \int_0^N \mathrm{d}n \phi_{pn} \left(\frac{k_\mathrm{s}}{\zeta} \frac{\partial^2 \boldsymbol{r}_n(t)}{\partial n^2} + \frac{1}{\zeta} \boldsymbol{f}_n(t) \right),\end{aligned}$$

where in the last step we have inserted the equation of motion eq. (A.1). Using integration by parts yields

$$\zeta_p \frac{\partial \boldsymbol{X}_p(t)}{\partial t} = \frac{k_\mathrm{s}\zeta_p}{\zeta} \phi_{pn} \frac{\partial \boldsymbol{r}_n(t)}{\partial n}\bigg|_0^N - \frac{k_\mathrm{s}\zeta_p}{\zeta} \int_0^N \mathrm{d}n \frac{\partial \phi_{pn}}{\partial n} \frac{\partial \boldsymbol{r}_n(t)}{\partial n} + \frac{\zeta_p}{\zeta} \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{f}_n(t)\,.$$

The first term vanishes according to the boundary conditions and integration by parts of the second term leads to

$$\zeta_p \frac{\partial \boldsymbol{X}_p(t)}{\partial t} = -\frac{k_\mathrm{s}\zeta_p}{\zeta} \frac{\partial \phi_{pn}}{\partial n} \boldsymbol{r}_n(t)\bigg|_0^N + \frac{k_\mathrm{s}\zeta_p}{\zeta} \int_0^N \mathrm{d}n \frac{\partial^2 \phi_{pn}}{\partial n^2} \boldsymbol{r}_n(t) + \frac{\zeta_p}{\zeta} \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{f}_n(t)\,.$$

Using the right hand side of the new Langevin equation it follows

$$\begin{aligned}-\frac{k_\mathrm{s}\zeta_p}{\zeta} \frac{\partial \phi_{pn}}{\partial n} \boldsymbol{r}_n(t)\bigg|_0^N + \frac{k_\mathrm{s}\zeta_p}{\zeta} \int_0^N \mathrm{d}n \frac{\partial^2 \phi_{pn}}{\partial n^2} \boldsymbol{r}_n(t) + \frac{\zeta_p}{\zeta} \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{f}_n(t) &\stackrel{!}{=} -k_p \boldsymbol{X}_p(t) + \boldsymbol{f}_p(t)\\ &= -k_p \int_0^N \mathrm{d}n \phi_{pn} \boldsymbol{r}_n(t) + \boldsymbol{f}_p(t)\,,\end{aligned}$$

where in the second step we again inserted the ansatz. For the equality to hold, ϕ_{pn} has to fulfil the following equations

$$\frac{k_s\zeta_p}{\zeta}\frac{\partial^2\phi_{pn}}{\partial n^2} = -k_p\phi_{pn} \tag{A.2}$$

$$\left.\frac{\partial\phi_{pn}}{\partial n}\right|_{n=0} = \left.\frac{\partial\phi_{pn}}{\partial n}\right|_{n=N} = 0 \tag{A.3}$$

$$\frac{\zeta_p}{\zeta}\int_0^N \mathrm{d}n\phi_{pn}\boldsymbol{f}_n(t) = \boldsymbol{f}_p(t)\,. \tag{A.4}$$

Equation (A.2) is the standard equation for oscillators with von Neumann boundary conditions eq. (A.3). From eq. (A.2) we make the ansatz $\phi_{pn} = A\cos(an+b)$. Using eq. (A.3) it follows $b = aN+b \overset{!}{=} \pi p$ and therefore $a = p\pi/N$ and $b = 0$, leading to $\phi_{pn} = A\cos(p\pi n/N)$. From eq. (A.2) it follows $k_p = k_s\zeta_p/\zeta(p\pi/N)^2$. The factor A is obtained from the normalisation condition [106] at $p = 0$, i.e. $\int_0^N \mathrm{d}nA\cos(p\pi n/N)|_{p=0} = AN \overset{!}{=} 1$ and therefore $\phi_{pn} = 1/N\cos(p\pi n/N)$. With this it follows

$$\boldsymbol{X}_p(t) = \frac{1}{N}\int_0^N \mathrm{d}n\cos\left(\frac{p\pi n}{N}\right)\boldsymbol{r}_n(t)\,,$$

with the inverse transform

$$\boldsymbol{r}_n(t) = \boldsymbol{X}_0(t) + 2\sum_{p=1}^{\infty}\boldsymbol{X}_p(t)\cos\left(\frac{p\pi n}{N}\right)\,.$$

The mean of the new random force vanishes by construction

$$\begin{aligned}\langle\boldsymbol{f}_p(t)\rangle &= \frac{\zeta_p}{\zeta}\int_0^N \mathrm{d}n\phi_{pn}\langle\boldsymbol{f}_n(t)\rangle\\ &= 0\,,\end{aligned}$$

and the second moment yields

$$\begin{aligned}\langle f_{\alpha p}(t)f_{\beta q}(t')\rangle &= \left\langle\frac{\zeta_p}{\zeta}\int_0^N \mathrm{d}n\phi_{pn}f_{\alpha n}(t)\frac{\zeta_q}{\zeta}\int_0^N \mathrm{d}m\phi_{qm}f_{\beta m}(t')\right\rangle\\
&= \frac{\zeta_p\zeta_q}{\zeta^2}\int_0^N \mathrm{d}n\int_0^N \mathrm{d}m\phi_{pn}\phi_{qm}\langle f_{\alpha n}(t)f_{\beta m}(t')\rangle\\
&= \frac{\zeta_p\zeta_q}{\zeta^2}\int_0^N \mathrm{d}n\int_0^N \mathrm{d}m\phi_{pn}\phi_{qm}2\zeta k_\mathrm{B}T\delta(n-m)\delta_{\alpha\beta}\delta(t-t')\\
&= \frac{\zeta_p\zeta_q}{N^2\zeta}2k_\mathrm{B}T\delta_{\alpha\beta}\delta(t-t')\int_0^N \mathrm{d}n\cos\left(\frac{p\pi n}{N}\right)\cos\left(\frac{q\pi n}{N}\right)\\
&= \frac{\zeta_p\zeta_q}{N^2\zeta}2k_\mathrm{B}T\delta_{\alpha\beta}\delta(t-t')\frac{N}{2\pi}\left(\frac{\sin\left(\frac{\pi n(p-q)}{N}\right)}{p-q} + \frac{\sin\left(\frac{\pi n(p+q)}{N}\right)}{p+q}\right)\Bigg|_0^N\end{aligned}$$

$$
\begin{aligned}
&= \frac{\zeta_p \zeta_q}{N^2 \zeta} 2 k_\mathrm{B} T \delta_{\alpha\beta} \delta(t - t') \frac{N}{2\pi} \delta(p - q) \pi (1 + \delta_{p0}) \\
&= \frac{\zeta_p^2}{N\zeta} k_\mathrm{B} T \delta_{\alpha\beta} \delta(t - t') (1 + \delta_{p0}) \,.
\end{aligned}
$$

Here we have used $\sin(\pi x)/x|_{x=0} = \pi$. We choose ζ_p such that the new random force fulfils the same equations as $\boldsymbol{f}_n$, i.e. $\langle f_{\alpha p}(t) f_{\beta q}(t') \rangle = 2\zeta_p k_\mathrm{B} T \delta_{\alpha\beta} \delta(p - q) \delta(t - t')$, from which we obtain

$$
2\zeta_p k_\mathrm{B} T = \frac{\zeta_p^2}{N\zeta} k_\mathrm{B} T (1 + \delta_{p0}) \,.
$$

Hence it follows $\zeta_p = 2N\zeta$ for $p > 0$ and $\zeta_p = N\zeta$ fpr $p = 0$ and $k_p = k_\mathrm{s} \zeta_p / \zeta (p\pi/N)^2 = 2k_\mathrm{s}/N(p\pi)^2$.

A2.2 Solving the discrete Rouse model

We follow the derivation of the normal modes shown in [107]. We aim at solving the Langevin equations describing the dynamics of a Rouse chain with $A := k_\mathrm{s}/\zeta$

$$
\begin{aligned}
\frac{\mathrm{d}\boldsymbol{r}_n(t)}{\mathrm{d}t} &= -A(2\boldsymbol{r}_n(t) - \boldsymbol{r}_{n+1}(t) - \boldsymbol{r}_{n-1}(t)) \quad \text{for } n = 1, \ldots, N-1 \\
\frac{\mathrm{d}\boldsymbol{r}_0(t)}{\mathrm{d}t} &= -A(\boldsymbol{r}_0(t) - \boldsymbol{r}_1(t)) \\
\frac{\mathrm{d}\boldsymbol{r}_N(t)}{\mathrm{d}t} &= -A(\boldsymbol{r}_N(t) - \boldsymbol{r}_{N-1}(t)) \,,
\end{aligned}
$$

by making the ansatz $\boldsymbol{r}_n(t) = \boldsymbol{X}(t) \cos(an + b)$ and attempting to solve the differential equations without the random forces. Inserting the ansatz yields

$$
\cos(b) \frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = -A\boldsymbol{X}(t) \left(\cos(b) - \cos(a + b)\right) \tag{A.5}
$$

$$
\cos(aN + b) \frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = -A\boldsymbol{X}(t) \left(\cos(aN + b) - \cos(a(N-1) + b)\right) \tag{A.6}
$$

$$
\begin{aligned}
\cos(an + b) \frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} &= -A\boldsymbol{X}(t) \left(2\cos(an + b) - \cos(a(n-1) + b) - \cos(a(n+1) + b)\right) \\
&= -A\boldsymbol{X}(t) \left(\cos(an + b)(2 - 2\cos(a))\right) \\
&= -A\boldsymbol{X}(t) \cos(an + b) 4 \sin^2\left(\frac{a}{2}\right) \,.
\end{aligned} \tag{A.7}
$$

The boundaries are consistent, when eq. (A.5) equals eq. (A.7) with $n = 0$:

$$
\begin{aligned}
\cos(b) - \cos(a + b) &= \cos(b) 4 \sin^2\left(\frac{a}{2}\right) \\
&= \cos(b)(2 - 2\cos(a)) \\
&= 2\cos(b) - \cos(a + b) - \cos(a - b) \\
0 &= \cos(b) - \cos(a - b) \,.
\end{aligned}
$$

From $\cos(b) = \cos(a-b)$ it follows $a = 2b$. Moreover, eq. (A.6) has to equal eq. (A.7) with $n = N$:

$$\begin{aligned}\cos(aN+b) - \cos(a(N-1)+b) &= \cos(aN+b)4\sin^2\left(\frac{a}{2}\right)\\ &= 2\cos(aN+b) - \cos(a(N-1)+b) - \cos(a(N+1)+b)\\ 0 &= \cos(aN+b) - \cos(a(N+1)+b)\\ &= \cos(aN+b) - \cos(-a(N+1)-b)\,.\end{aligned}$$

From $\cos(aN+b) = \cos(-a(N+1)-b)$ it follows $aN+b = 2\pi p - a(N+1) - b$ and therefore $2\pi p = (2N+1)a + 2b = (2N+1)a + a = 2(N+1)a$ so that $a = \pi p/(N+1)$ and $b = \pi p/(2(N+1))$. The solution is the linear combination of all independent solutions, where the factor two is inserted for convenience [107]

$$\boldsymbol{r}_n(t) = \boldsymbol{X}_0 + 2\sum_{p=1}^{N} \boldsymbol{X}_p \cos\left(\frac{p\pi}{N+1}\left(n+\frac{1}{2}\right)\right)\,.$$

This is a discrete cosine transform (DCT-3), the inverse of which is [207]

$$\boldsymbol{X}_p = \frac{1}{N+1}\sum_{n=0}^{N} \boldsymbol{r}_n \cos\left(\frac{p\pi}{N+1}\left(n+\frac{1}{2}\right)\right)\,,$$

which is obtained using [107]

$$\frac{1}{N+1}\sum_{n=0}^{N} \cos\left(\frac{p\pi}{N+1}\left(n+\frac{1}{2}\right)\right) = \delta_{p0} \quad \text{for } 0 \geq p < 2(N+1)\,.$$

The equations of motion then read

$$\frac{\mathrm{d}\boldsymbol{X}_p}{\mathrm{d}t} = -\frac{k_\mathrm{s}}{\zeta} 4\sin^2\left(\frac{p\pi}{2(N+1)}\right)\boldsymbol{X}_p + \boldsymbol{f}_p\,,$$

with the new random force fulfilling

$$\begin{aligned}\langle \boldsymbol{f}_p(t)\rangle &= 0\\ \langle \boldsymbol{f}_0(t)\boldsymbol{f}_0(t')\rangle &= \frac{2D}{N+1}\delta(t-t')\\ \langle \boldsymbol{f}_p(t)\boldsymbol{f}_q(t')\rangle &= \frac{D}{N+1}\delta_{pq}\delta(t-t') \quad \text{for } p+q>0\,.\end{aligned}$$

A3 Centre of Mass Diffusion within the Rouse Model

To calculate the mean squared end-to-end distance of the centre of mass $g_3 = \langle(\boldsymbol{R}_\mathrm{cm}(t) - \boldsymbol{R}_\mathrm{cm}(0))^2\rangle$, we insert the definition of the centre of mass $\boldsymbol{R}_\mathrm{cm}(t) = \frac{1}{(N+1)}\sum_{n=0}^{N}\boldsymbol{r}_n(t)$ for a Rouse chain of $N+1$ beads into the Langevin equation of the Rouse model:

$$\zeta\frac{\mathrm{d}\boldsymbol{R}_\mathrm{cm}(t)}{\mathrm{d}t} = -\frac{k_\mathrm{s}}{N+1}\left(\boldsymbol{r}_0 - \boldsymbol{r}_1 + \sum_{n=1}^{N-1}(2\boldsymbol{r}_n - \boldsymbol{r}_{n+1} - \boldsymbol{r}_{n-1}) + \boldsymbol{r}_N - \boldsymbol{r}_{N-1}\right) + \frac{1}{N+1}\sum_{n=0}^{N}\boldsymbol{f}_n(t)$$

$$= \frac{1}{N+1} \sum_{n=0}^{N} \boldsymbol{f}_n(t)\,,$$

since the first sum vanishes. Integrating both sides gives

$$\int \mathrm{d}\boldsymbol{R}_{\mathrm{cm}}(t) = \frac{1}{\zeta(N+1)} \int \sum_{n=0}^{N} \boldsymbol{f}_n(t)\mathrm{d}t'$$

$$\boldsymbol{R}_{\mathrm{cm}}(t) - \boldsymbol{R}_{\mathrm{cm}}(0) = \frac{1}{\zeta(N+1)} \int_0^t \sum_{n=0}^{N} \boldsymbol{f}_n(t)\mathrm{d}t'\,,$$

and therefore the mean squared end-to-end distance of the centre of mass is

$$\begin{aligned}
\langle (\boldsymbol{R}_{\mathrm{cm}}(t) - \boldsymbol{R}_{\mathrm{cm}}(0))^2 \rangle &= \left\langle \left(\frac{1}{\zeta(N+1)} \int_0^t \sum_{n=0}^{N} \boldsymbol{f}_n(t)\mathrm{d}t' \right)^2 \right\rangle \\
&= \frac{1}{\zeta^2(N+1)^2} \left\langle \int_0^t \sum_{n=0}^{N} \boldsymbol{f}_n(t')\mathrm{d}t' \int_0^t \sum_{m=0}^{N} \boldsymbol{f}_m(s')\mathrm{d}s' \right\rangle \\
&= \frac{1}{\zeta^2(N+1)^2} \int_0^t \mathrm{d}t' \int_0^t \mathrm{d}s' \sum_{n,m=0}^{N} \langle \boldsymbol{f}_n(t')\boldsymbol{f}_m(s') \rangle \\
&= \frac{1}{\zeta^2(N+1)^2} \int_0^t \mathrm{d}t' \int_0^t \mathrm{d}s' \sum_{n,m=0}^{N} 6\zeta k_{\mathrm{B}} T \delta_{nm} \delta(t'-s') \\
&= \frac{1}{\zeta(N+1)^2} \int_0^t \mathrm{d}t' \sum_{n=0}^{N} 6k_{\mathrm{B}}T \\
&= \frac{1}{\zeta(N+1)} 6k_{\mathrm{B}}Tt = 6D_{\mathrm{cm}}t\,.
\end{aligned}$$

We used the known relations $\langle f_{n\alpha}(t) f_{m\beta}(t') \rangle = 2\zeta k_{\mathrm{B}} T \delta_{nm} \delta_{\alpha\beta} \delta(t-t')$ as well as $D_{\mathrm{cm}} = k_{\mathrm{B}}T/((N+1)\zeta)$.

Testing and Validation of Different Polymer Dynamics Models

B

B1 Evaluation of Brownian Dynamics Simulations of the Rouse Model

For the BD simulations of a Rouse chain the spring potential eq. (II.2) is used with an entropic spring constant $k_s = 3k_BT/l_K^2$, with equilibrium distance $l_0 = 0$ and no bending potential $U_b = 0$. Lengths are measured in terms of the Kuhn length and energies in units of k_BT. The dimensionless parameters are $\boldsymbol{r}' = \boldsymbol{r}/l_K$, $U' = U/(k_BT)$, $k_s' = k_s l_K^2/(k_BT)$, $\boldsymbol{F}' = \boldsymbol{F} l_K/(k_BT)$ and $\Delta t' = \Delta t \mu_b k_B T/l_K^2$. With this the dimensionless form of the discretised Langevin equation eq. (II.1) becomes, using $\boldsymbol{F}_i(t) = -\boldsymbol{\nabla}_{\boldsymbol{r}_i} U$

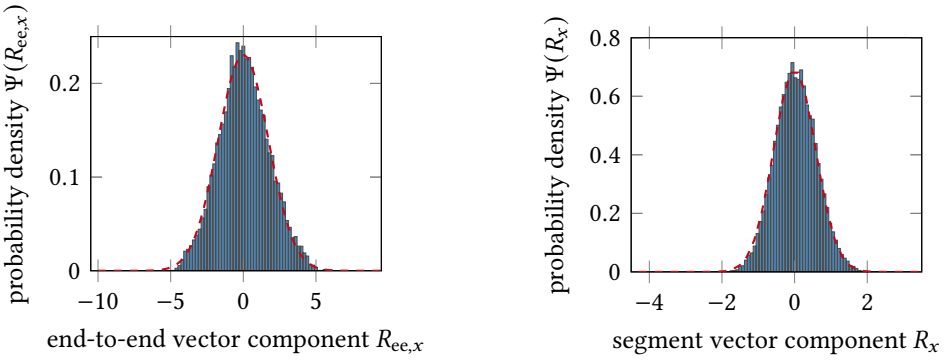
$$\begin{aligned}
\boldsymbol{r}_i'(t+\Delta t) &= \frac{\boldsymbol{r}_i(t+\Delta t)}{l_K} \\
&= \frac{1}{l_K}\left(\boldsymbol{r}_i(t) + \mu_b \Delta t \boldsymbol{F}_i(t) + \sqrt{2\mu_b k_B T \Delta t}\boldsymbol{\eta}_i(t)\right) \\
&= \boldsymbol{r}_i'(t) + \frac{\mu_b \Delta t \boldsymbol{F}_i(t)}{l_K}\frac{k_B T l_K}{k_B T l_K} + \sqrt{\frac{2\mu_b k_B T \Delta t}{l_K^2}}\boldsymbol{\eta}_i(t) \\
&= \boldsymbol{r}_i'(t) + \Delta t' \boldsymbol{F}_i'(t) + \sqrt{2\Delta t'}\boldsymbol{\eta}_i(t)\,.
\end{aligned}$$

With $k_B = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$, $T = 298\,\mathrm{K}$ and $l_K = 100\,\mathrm{nm}$ for dsDNA the spring constant is computed. Setting $\mu_b' = 1$ yields $D' = 1$, such that the simulations can be rescaled afterwards with an appropriate bead diffusion constant of the respective polymer. The parameters are summarised in table B.1.

Parameter	Real value	Simulation
k_s	$1.2300 \times 10^{-6}\,\mathrm{N\,m^{-1}}$	3
U_b	$0\,\mathrm{N\,m}$	0
l_k	$100\,\mathrm{nm}$	1
l_0	$0\,\mathrm{nm}$	0

Table B.1: Parameter set for the Rouse model.

The time step $\Delta t'$ has to be chosen with care, since large values may lead to inaccurate results, while unnecessarily small values waste CPU time. We therefore vary the time step and verify that the simulated chain exhibits the properties of a Gaussian chain and reproduces Rouse dynamics.



(a) end-to-end vector $R_{ee,x}$ **(b)** segment vector R_x

Figure B.1: Gaussian distribution of the segment vectors and the end-to-end vector. Results from BD simulations of a free Rouse chain. (**a**) Probability distribution of the x-component of the end-to-end vector $\boldsymbol{R}_{ee}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = N/3$ (red). (**b**) Probability distribution of the x-component of the segment vector $\boldsymbol{R}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1/3$ (red).

In the Gaussian chain model the segments $\boldsymbol{R}$ are normally distributed according to eq. (I.7). Inserting the average segment length l_K this yields

$$\Psi(\boldsymbol{R}) = \left(\frac{3}{2\pi l_K^2}\right)^{\frac{3}{2}} \exp\left[-\frac{3\boldsymbol{R}^2}{2l_K^2}\right],$$

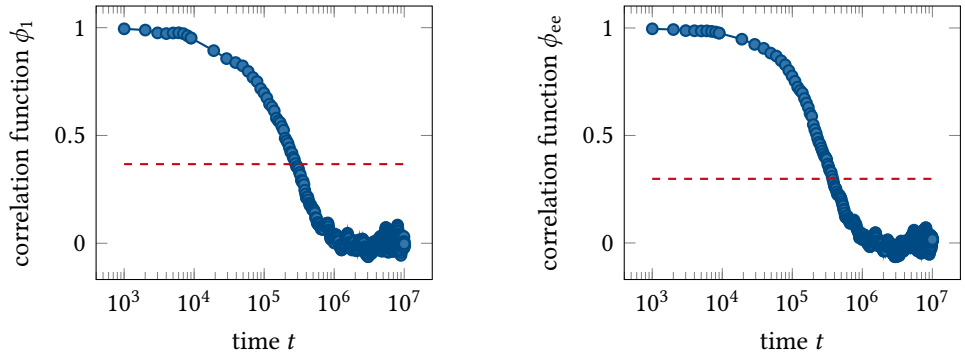
with the mean squared segment length $\langle \boldsymbol{R}^2 \rangle = l_K^2$. Therefore the x-components R_x of the segment vectors are distributed according to

$$\Psi(R_x) = \left(\frac{3}{2\pi l_K^2}\right)^{\frac{1}{2}} \exp\left[-\frac{3R_x^2}{2l_K^2}\right],$$

with $l_K' = 1$ in the simulations, so that the x-component is normally distributed with mean $\mu = 0$ and variance $\sigma^2 = 1/3$. From the distribution of the end-to-end vector in eq. (I.5) the distribution of the x-component yields

$$\Psi(R_{ee,x}) = \left(\frac{3}{2\pi N l_K^2}\right)^{\frac{1}{2}} \exp\left[-\frac{3R_{ee,x}^2}{2N l_K^2}\right].$$

In table B.2 the mean and variance of the segment length are shown for varying the size of the time step in the simulation. The correct values are obtained for time steps $\Delta t' = 0.001$ and smaller. We choose a time step of $\Delta t' = 0.00001$ in order to obtain the dynamics of the polymer even at very small times. In fig. B.1 the distribution of the x-value of the polymer segments and of the end-to-end vector is depicted for the chosen time step and a polymer of length $N = 10$. Both are shown to agree very well with the expected normal distribution.



(a) Correlation of the first Rouse mode

(b) End-to-end vector correlation

Figure B.2: Correlation functions. Results from BD simulations of a free Rouse chain. (**a**) Correlation function of the first Rouse mode ϕ_1 as a function of time. The dashed line represents $\phi_1(\tau_R) \approx 0.367$. (**b**) Correlation function of the end-to-end vector ϕ_{ee}. The dashed line represents $\phi_{ee}(\tau_R) \approx 0.298221$.

$\Delta t'$	$\langle \boldsymbol{R}'_x \rangle$	$\langle (\boldsymbol{R}'_x - \langle \boldsymbol{R}'_x \rangle)^2 \rangle$
0.1	−0.0012	0.5229
0.01	0.0008	0.3429
0.001	0.0020	0.3347
0.0001	−0.0066	0.3319
0.00001	0.0140	0.3310
0.000001	0.0044	0.3361
0 (exact)	0.0000	0.3333

Table B.2: Varying the time step for the Rouse model.

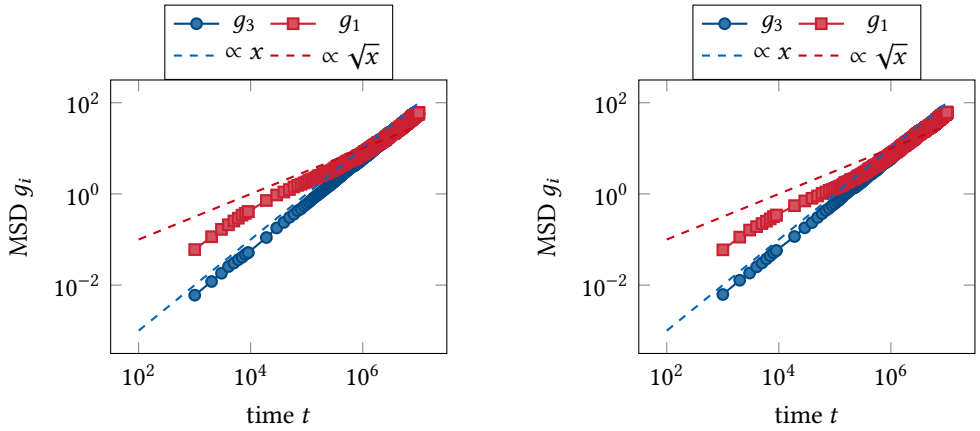
To validate the dynamics, we compute the dimensionless Rouse time. From eq. (I.15) with the average bond length l_K we obtain for a chain of N beads

$$\begin{aligned}\tau_R &= \frac{\zeta N^2 l_K^2}{k_B T 3\pi^2} \\ &= \frac{\zeta N^2 l_K^2}{D 3\pi^2} .\end{aligned}$$

Using $\Delta t' = \Delta t D / l_K^2$ yields

$$\begin{aligned}\tau'_R &= \frac{\zeta N^2 l_K^2}{D 3\pi^2} \frac{D}{\Delta t' l_K^2} \\ &= \frac{N^2}{3\pi^2 \Delta t'} .\end{aligned}$$

With the number of beads $N = 10$ and $\Delta t' = 0.00001$ this yields $\tau'_R \approx 337737$. For comparison with our simulations we compute the correlation functions. The correlation function of the



(a) Random initial configuration

(b) Linear initial configuration

Figure B.3: BD simulations of the Rouse model. The MSD of the centre of mass diffusion g_3 and the MSD of the central monomer g_1 are shown as a function of time t for BD simulations of a free Rouse polymer. (**a**) The polymer is initialised with a random walk. (**b**) The initial polymer configuration is a straight line. The dashed lines show the scaling of g_1 and g_3 as expected from the Rouse model.

pth Rouse mode is given by [208]

$$\begin{aligned}\phi_p(t) &= \frac{\langle \boldsymbol{X}_p(t)\boldsymbol{X}_p(0)\rangle}{\langle \boldsymbol{X}_p(0)^2\rangle} \\ &= \exp\left[\frac{-tp^2}{\tau_\mathrm{R}}\right],\end{aligned}$$

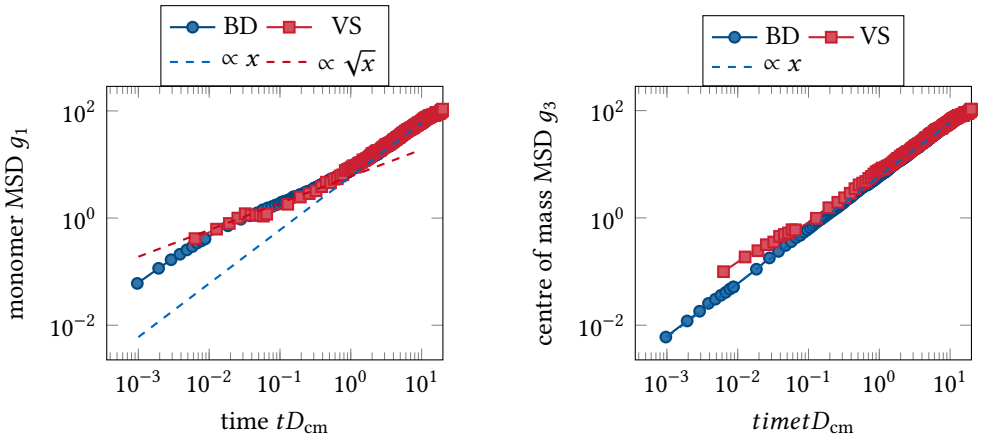
The dimensionless Rouse time can therefore be obtained from $\phi_1(\tau'_\mathrm{R}) = e^{-1} \approx 0.367$. The correlation function of the first Rouse mode, depicted in fig. B.2a first reaches a value of $\phi_1 = 0.367$ for $t' \approx 320649$ which is in quite good agreement with $\tau'_\mathrm{R} = 337737$. The correlation function of the end-to-end vector is [208]

$$\begin{aligned}\phi_\mathrm{ee}(t) &= \frac{\langle \boldsymbol{R}_\mathrm{ee}(t)\boldsymbol{R}_\mathrm{ee}(0)\rangle}{\langle \boldsymbol{R}_\mathrm{ee}(0)^2\rangle} \\ &= \sum_{p=1\ \mathrm{odd}}^{\infty} \frac{8}{p^2\pi^2} \exp\left[\frac{-tp^2}{\tau_\mathrm{R}}\right],\end{aligned}$$

such that $\phi_\mathrm{ee}(\tau_\mathrm{R}) \approx 0.298221$. The correlation function of the end-to-end vector, depicted in fig. B.2b first reaches a value of $\phi_\mathrm{ee} = 0.298221$ for $t' \approx 337683$ which is in very good agreement with $\tau'_\mathrm{R} = 337737$. Moreover, we compute the mean squared displacement of the centre of mass $g_3(t)$, as well as the mean squared displacement of the central monomer $g_1(t)$ and compare it to the scaling expected from the Rouse model shown in eq. (I.16) and eq. (I.17). In fig. B.3 the results are shown for simulations of a free polymer chain with random initial configuration (fig. B.3a) and with linear initial configuration (fig. B.3b) respectively. Both

plots show the expected linear scaling with time of the centre of mass displacement and the two scaling regimes of the monomer displacement. Our parameter choice for the BD simulations of the Rouse model therefore reproduces successfully both static and dynamic properties of the model.

B2 Evaluation of the Verdier-Stockmayer Move Set



(a) Monomer displacement

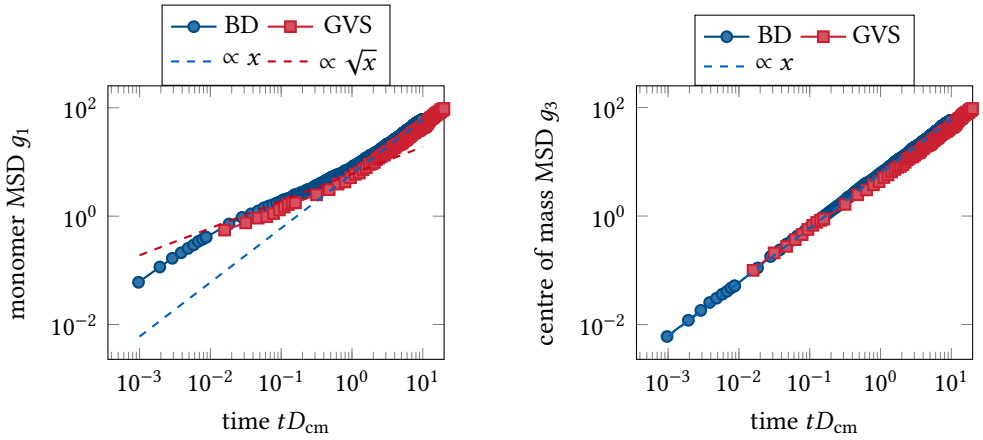
(b) Centre of mass displacement

Figure B.4: Verdier-Stockmayer move set with random initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a randomly initialised polymer simulated with the Verdier-Stockmayer move set (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. The dashed lines show the scaling of g_1 as expected from the Rouse model. (**b**) The corresponding MSD of the centre of mass g_3.

To validate the dynamics of a polymer simulated with the Verdier-Stockmayer move set consisting of an end-move and a kink-jump, the diffusion of a free polymer is simulated and compared to the BD simulations of the Rouse model. In fig. B.4 the resulting mean squared displacement of the central monomer g_1 and the mean squared displacement of the centre of mass g_3 are compared to the BD simulations and the expected scaling according to the Rouse model. The chain in the concerning plots has been initialised as a random walk on the lattice and the scaling of g_1 and g_3 agrees very well with the Rouse model.

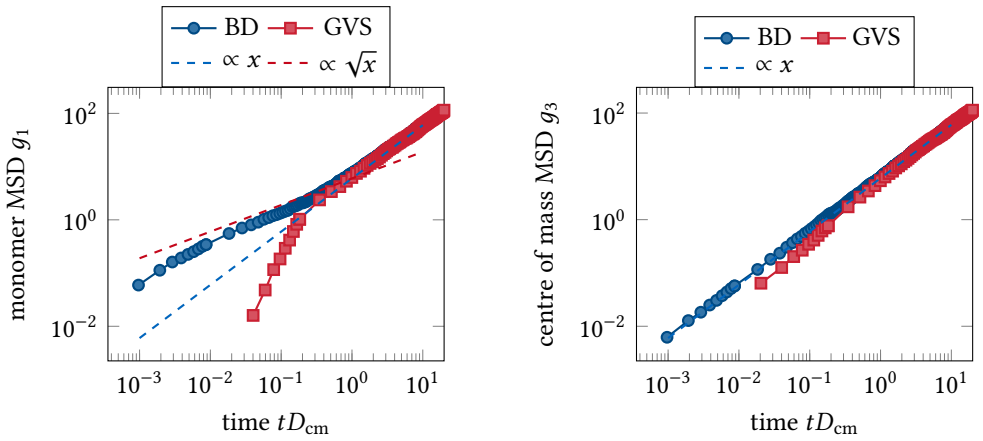
Figure II.4 shows the corresponding plots for a polymer that was initialised in a linear configuration on the lattice. Since for such a configuration the inner chain segments are essentially immobile until the movement has diffused from the chain ends towards the inner segments, the computed displacement of the central monomer deviates strongly from the expected Rouse scaling at short times.

In fig. B.5 the mean squared displacement of the centre of mass and the mean squared displacement of the central monomer computed from simulations of a free polymer with the



(a) Monomer displacement **(b)** Centre of mass displacement

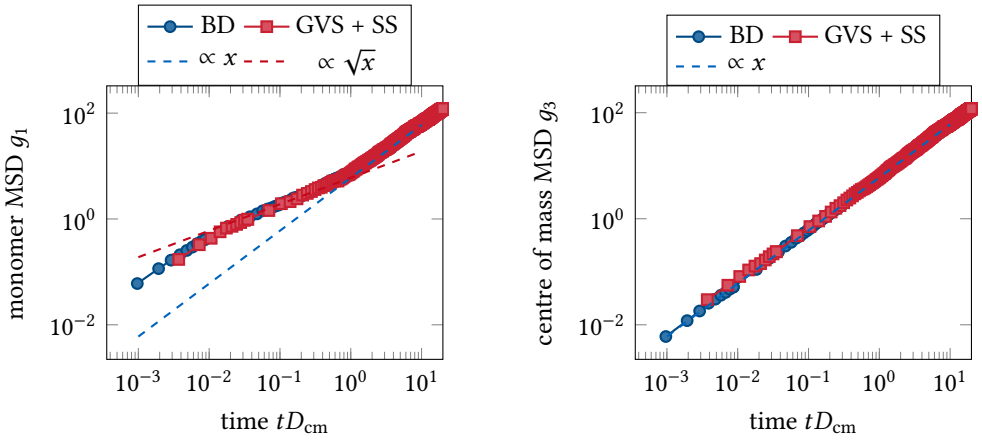
Figure B.5: Generalised Verdier-Stockmayer move set with random initial configuration. The MSD of the central monomer g_1 as a function of time for a randomly initialised polymer simulated with the generalised Verdier-Stockmayer move set (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{\mathrm{cm}}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.



(a) Monomer displacement **(b)** Centre of mass displacement

Figure B.6: Generalised Verdier-Stockmayer move set with linear initial configuration. The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the generalised Verdier-Stockmayer move set (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{\mathrm{cm}}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

generalised Verdier-Stockmayer move set are shown in comparison to BD simulations of the



(a) Monomer displacement

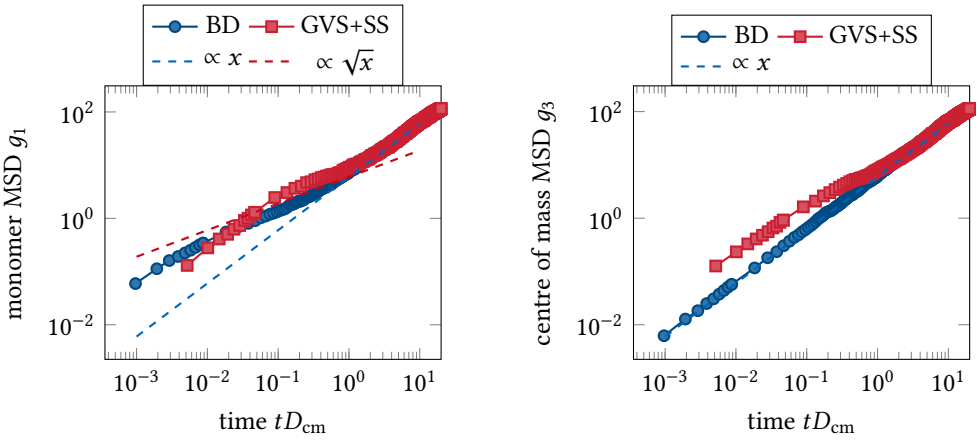
(b) Centre of mass displacement

Figure B.7: Generalised Verdier-Stockmayer move set and slithering snake with random initial configuration. The MSD of the central monomer g_1 as a function of time for a polymer with random initial configuration simulated with the generalised Verdier-Stockmayer move set and the slithering snake move (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

Rouse model. The chain was initialised as a random walk and shows very good agreement with the Rouse model, as expected.

Adding the crankshaft move to the Verdier-Stockmayer move set however does not resolve the slow relaxation of an initially linear chain, as demonstrated in fig. B.6. The crankshaft move allows the chain to move out of the plane but starting from a linear configuration the movement still has to diffuse from the chain ends towards the inner segments.

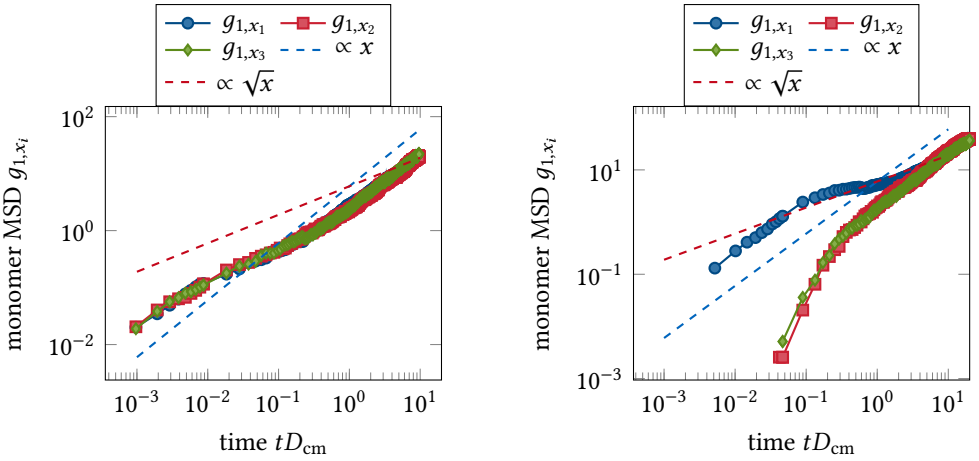
Adding a slithering snake move to the generalised Verdier-Stockmayer move set still reproduces Rouse dynamics when the polymer is initialised as a random walk, as demonstrated in fig. B.7. Also for an initially linear chain the dynamics of the MSD of the central monomer show a better agreement with the Rouse model than without the slithering snake move, shown in fig. B.8a. However, the MSD of the centre of mass displays some deviation from the expected scaling, see fig. B.8b. This is because the slithering snake move leads to faster relaxation of the central monomer displacement only in the direction in which the chain can reptate. This is demonstrated in fig. B.9, where the different components g_{1,x_i} of the MSD of the central monomer in direction x_i for $i \in \{1, 2, 3\}$ are shown as a function of time. BD simulations of the Rouse model with an initially linear chain, depicted in fig. B.9a, show that for all three components the usual Rouse scaling is expected. From simulations of the Verdier-Stockmayer move set with the slithering snake move however, shown in fig. B.9b, the MSD in the $x-$direction in which the chain can reptate the relaxation is much faster, while the dynamics of the two other directions are still as slow as without the slithering snake move.



(a) Monomer displacement

(b) Centre of mass displacement

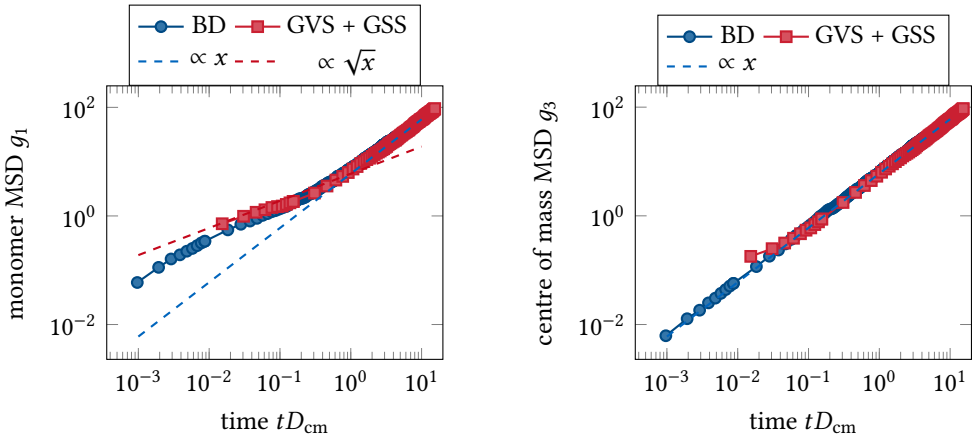
Figure B.8: Generalised Verdier-Stockmayer move set and slithering snake with linear initial configuration. The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the generalised Verdier-Stockmayer move set and the slithering snake move (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.



(a) BD simulation

(b) Generalised Verdier-Stockmayer with slithering snake

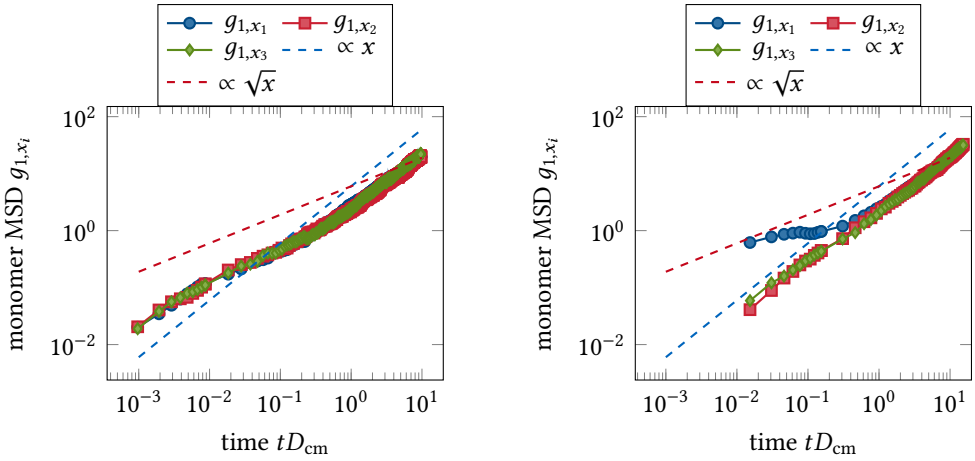
Figure B.9: Components of the monomer displacement. (**a**) The MSD of the central monomer g_{1,x_i} in spatial direction x_i is shown as a function of time for a polymer with linear initial configuration from BD simulations of the Rouse model. The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. (**b**) The corresponding plot simulated with the generalised Verdier-Stockmayer move set and the slithering snake move.



(a) Monomer displacement

(b) Centre of mass displacement

Figure B.10: Generalised Verdier-Stockmayer move set and generalised slithering snake with linear initial configuration. The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the generalised Verdier-Stockmayer move set and the generalised slithering snake move (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.



(a) BD simulations

(b) Generalised Verdier-Stockmayer and generalised slithering snake

Figure B.11: Components of the monomer displacement. (**a**) The MSD of the central monomer g_{1,x_i} in spatial direction x_i is shown as a function of time for a polymer with linear initial configuration from BD simulations of the Rouse model. The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. (**b**) The corresponding plot simulated with the generalised Verdier-Stockmayer move set and the generalised slithering snake move.

To yield robust Rouse dynamics independent of the initial configuration of the chain even more moves would have to be added. As mentioned in the main text a generalised slithering snake move was constructed to resolve the problem that a linear chain can only reptate in one direction. As shown in fig. B.10 both the monomer displacement and the centre of mass displacement of an initially linear polymer simulated with the generalised Verdier-Stockmayer move set and the generalised slithering snake move agree quite well with the Rouse model. However, as demonstrated in fig. B.11 the generalised slithering snake move still leads to strong deviations of the spatial components of the monomer displacement from the Rouse model. For an initially linear chain the relaxation in $x-$direction is still too fast, while there is too little movement in the two other directions.

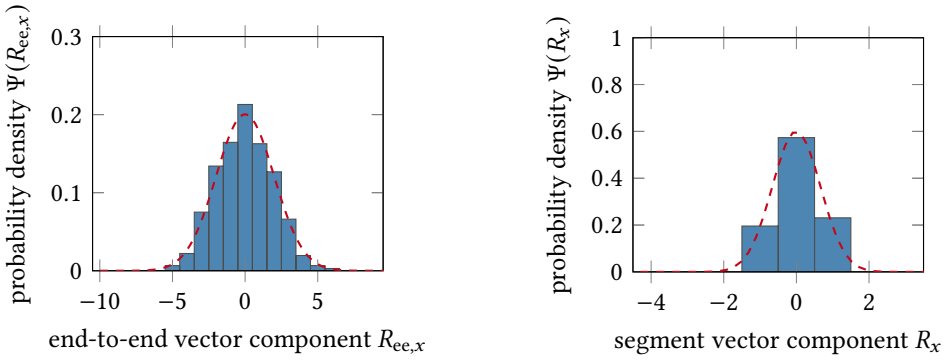
B3 Evaluation of the Single-Site Bond-Fluctuation Model with Fraenkel Spring

To simulate a FJC on a cubic lattice using a ssBFM with Fraenkel springs connecting the beads, the spring has to be chosen stiff enough to reproduce the properties of a FJC but soft enough to exhibit Rouse dynamics. For different spring constants k_s' the average values of the segment length R, the end-to-end vector $\boldsymbol{R}_{ee}$ and the angle between neighbouring segments are computed, as shown in table B.3. For all shown spring constants a good agreement with the FJC model is obtained. To further investigate the properties of the lattice simulations, the distribution of the end-to-end vector and of the segment vector are analysed. For a spring constant of $k_s' = 10$ the distribution of the x-component of the end-to-end vector and of the segment vector are shown in fig. B.12 and compared to the expected Gaussian distribution, yielding quite a good agreement.

$\boldsymbol{k_s'}$	$\langle \boldsymbol{R'} \rangle$	$\langle (\boldsymbol{R'} - \langle \boldsymbol{R'} \rangle)^2 \rangle$	$\langle \boldsymbol{R_{ee}^2} \rangle$	$\langle \cos \rangle$	$\langle (\cos - \langle \cos \rangle)^2 \rangle$
10	1.1310	0.0386	11.3520	−0.0095	0.3279
20	1.0593	0.0211	10.3944	0.0056	0.3270
30	1.0426	0.0158	9.9682	−0.0006	0.3280
50	1.0390	0.0146	10.0593	0.0005	0.3238
100	1.0388	0.0146	10.1190	0.0025	0.3261
exact	1	0	9	0	0.3300

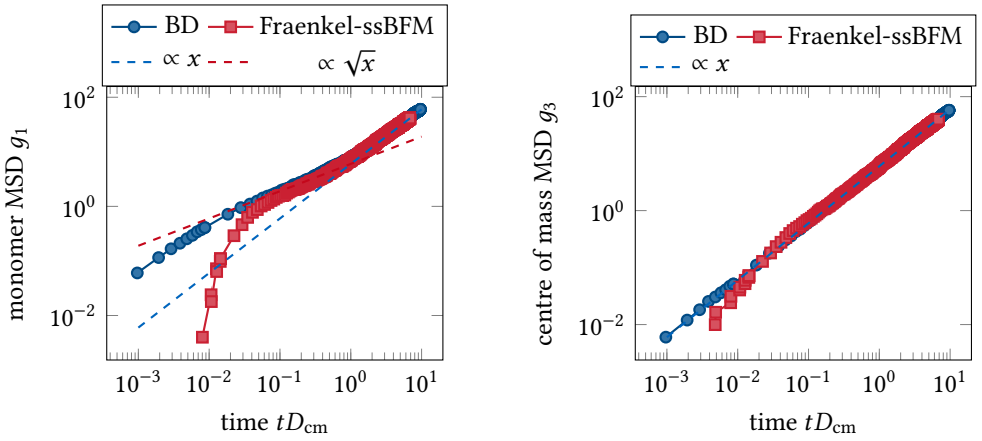
Table B.3: Varying the spring constant for the FJC lattice model.

For a spring constant $k_s' = 30$ the mean squared displacement of the centre of mass and of the central monomer are computed and compared to BD simulations of a Rouse chain. As depicted in fig. B.13 a randomly initialised polymer exhibits some deviations for small times but still shows the expected Rouse scaling at larger times. For a polymer that is initially in a linear configuration however there is no subdiffusive regime of the monomer displacement g_1 as expected from the Rouse model. A spring constant of $k_s' = 30$ being apparently too stiff to reproduce stable Rouse dynamics, we therefore soften the spring until we obtain a satisfactory result.



(a) end-to-end vector $R_{ee,x}$ **(b)** segment vector R_x

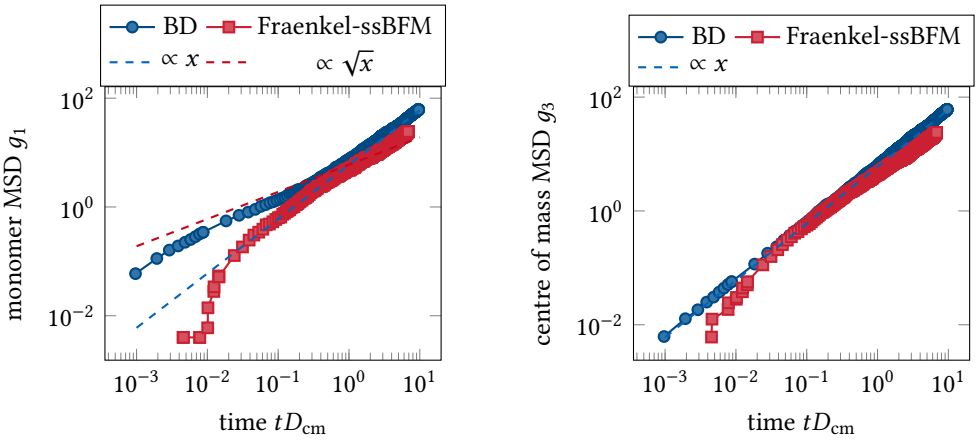
Figure B.12: Gaussian distribution of the segment vectors and the end-to-end vector. (**a**) Probability distribution of the x-component of the end-to-end vector $\boldsymbol{R}_{ee}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = \langle b^2 \rangle N/3$ (red). (**b**) Probability distribution of the x-component of the segment vector $\boldsymbol{R}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = \langle b^2 \rangle / 3$ (red) with $\langle b^2 \rangle = 1.32$. The Fraenkel spring constant is $k_s' = 10$.



(a) Monomer displacement **(b)** Centre of mass displacement

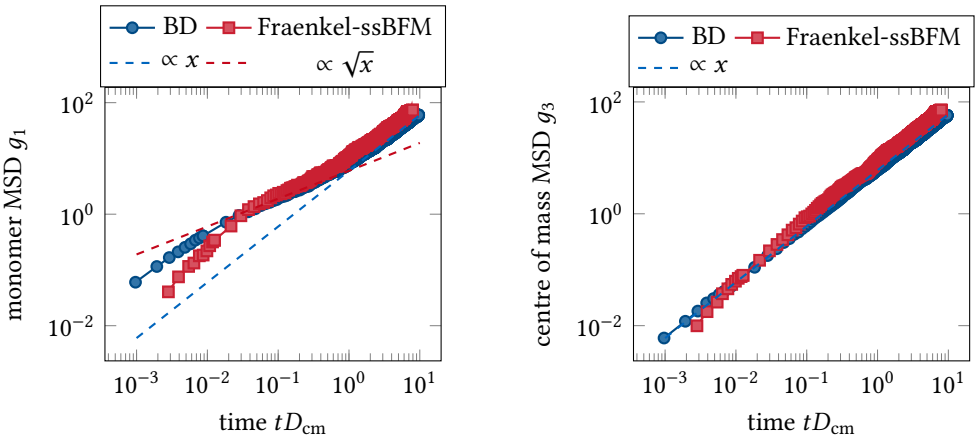
Figure B.13: Fraenkel-ssBFM with random initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a randomly initialised polymer simulated with the ssBFM with Fraenkel spring $k_s' = 30$ (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

Figure B.15 and fig. B.16 show the same simulations with a spring constant $k_s' = 10$, which shows a quite stable Rouse scaling even when the polymer is initially in a linear configuration. We therefore choose a spring constant of $k_s' = 10$, which yields polymer configurations that



(a) Monomer displacement **(b)** Centre of mass displacement

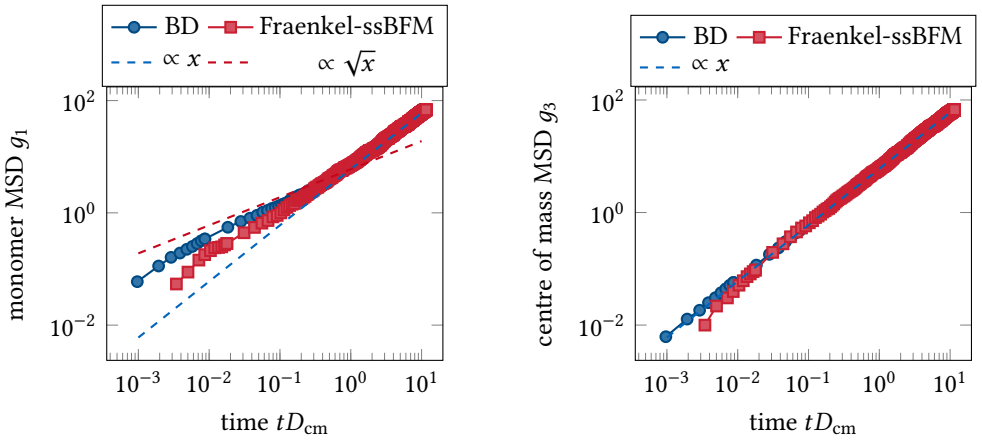
Figure B.14: Fraenkel-ssBFM with linear initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the ssBFM with Fraenkel spring $k_s' = 30$ (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.



(a) Monomer displacement **(b)** Centre of mass displacement

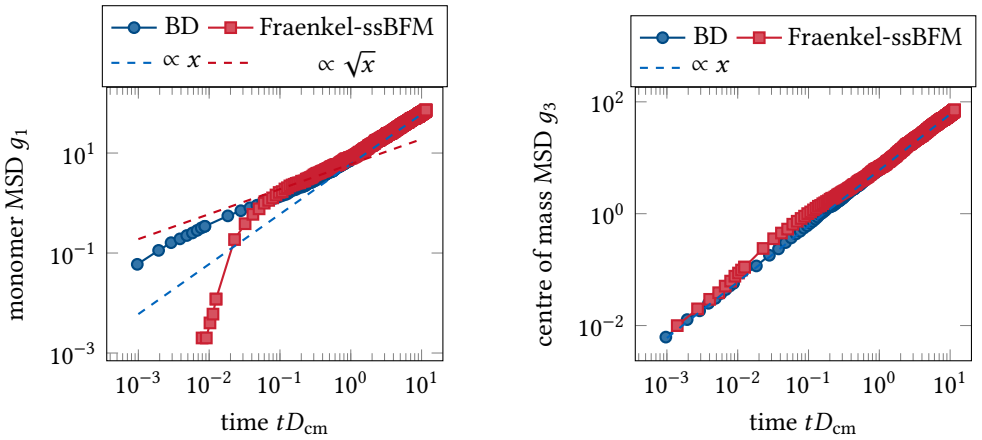
Figure B.15: Fraenkel-ssBFM with random initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a randomly initialised polymer simulated with the ssBFM with Fraenkel spring $k_s' = 10$ (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

agree with the FJC sufficiently well but is also soft enough to reproduce stable Rouse dynamics even when starting from a linear initial chain configuration.



(a) Monomer displacement **(b)** Centre of mass displacement

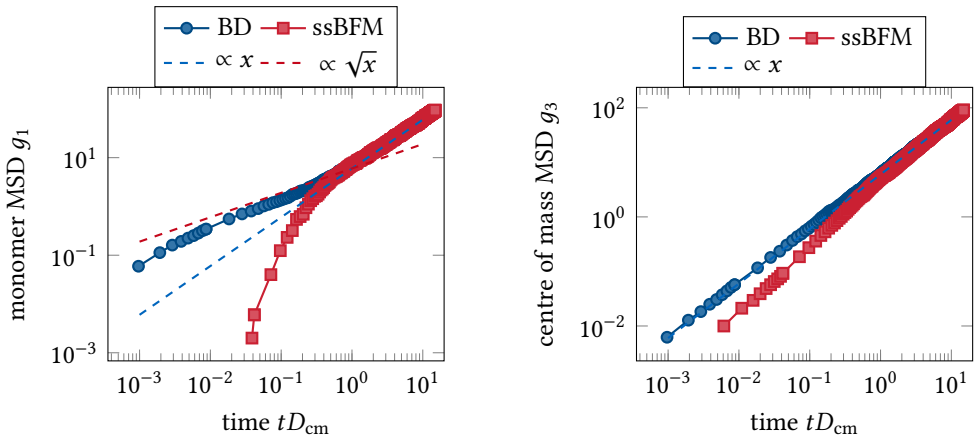
Figure B.16: Fraenkel-ssBFM with linear initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the ssBFM with Fraenkel spring $k_s' = 10$ (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.



(a) Monomer displacement **(b)** Centre of mass displacement

Figure B.17: Fraenkel-ssBFM with stretched linear initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a polymer with stretched linear initial configuration (bead distance $\sqrt{3}$) simulated with the ssBFM with Fraenkel spring $k_s' = 10$ (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

B4 Evaluation of the Phantom Chain Single-Site Bond-Fluctuation Model



(a) Monomer displacement

(b) Centre of mass displacement

Figure B.18: Phantom chain ssBFM with with stretched linear initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a polymer with stretched linear initial configuration (bead distance $\sqrt{3}$) simulated with the ssBFM (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{\mathrm{cm}}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

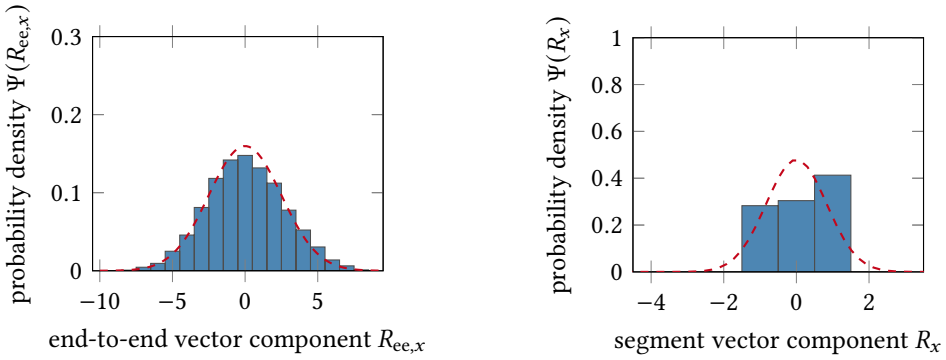
To evaluate the phantom chain implementation of the single-site Bond-Fluctuation model, we first compute some expectation values of the model. From the three different lengths of the segments $b \in \{1, \sqrt{2}, \sqrt{3}\}$ with probabilities $\{3/13, 6/13, 4/13\}$ the average segment length yields

$$\begin{aligned}\langle R\rangle &= \sum_i R_i p_i \\ &= \frac{1}{13}(3 + 6\sqrt{2} + 4\sqrt{3}) \approx 1.416\,,\end{aligned}$$

and the average squared segment length

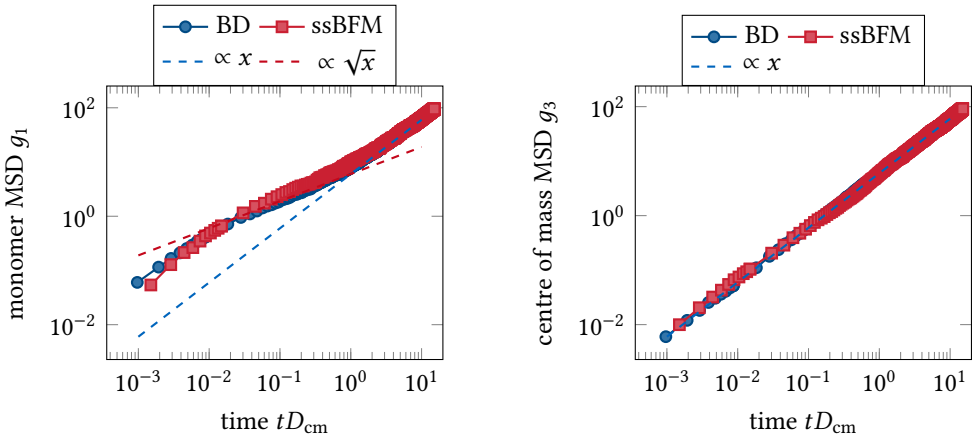
$$\langle R^2\rangle = \frac{1}{13}(3 + 12 + 12) \approx 2.0769\,.$$

In table B.4 the average values of the segment length, the end-to-end vector and the angle between neighbouring segments are compared to the expected values for the FJC. The computed values are in very good agreement. To further investigate the lattice model the end-to-end vector and the segment vectors are analysed. The distribution of the x-component of the end-to-end vector and the segment vector are depicted in fig. B.19 and compared to the Gaussian distribution expected from the FJC model, yielding quite a good agreement.



(a) end-to-end vector $R_{ee,x}$ **(b)** segment vector R_x

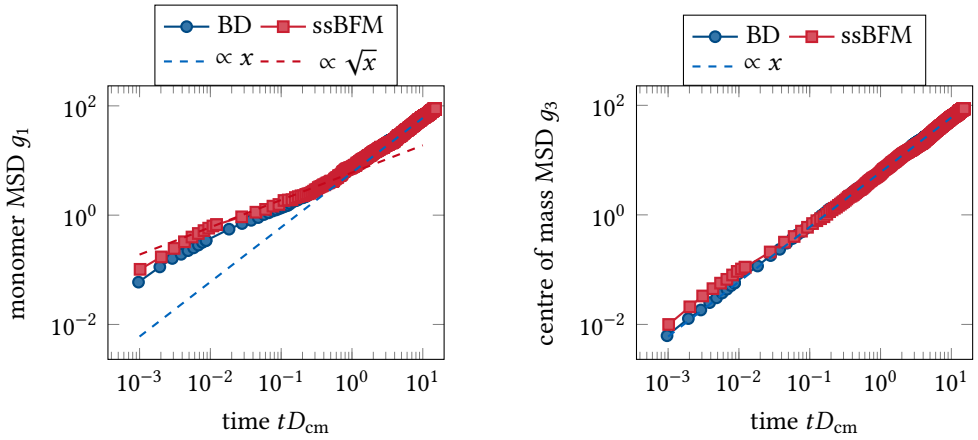
Figure B.19: Gaussian distribution of the segment vectors and the end-to-end vector. (**a**) Probability distribution of the x-component of the end-to-end vector $\boldsymbol{R}_{ee}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = \langle b^2 \rangle N/3$ (red). (**b**) Probability distribution of the x-component of the segment vector $\boldsymbol{R}$ (blue) in comparison to the Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = \langle b^2 \rangle / 3$ (red) with $\langle b^2 \rangle = 2.08$.



(a) Monomer displacement **(b)** Centre of mass displacement

Figure B.20: Phantom chain ssBFM with random initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a randomly initialised polymer simulated with the ssBFM (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

Furthermore, the Rouse scaling of the phantom chain ssBFM is analysed. In figs. B.20 and B.21 the mean squared displacement of the centre of mass and of the central monomer are compared to BD simulations of the Rouse model for a random initial configuration and a linear initial chain configuration respectively. Both plots show a very good agreement,



(a) Monomer displacement

(b) Centre of mass displacement

Figure B.21: Phantom chain ssBFM with linear initial configuration. (**a**) The MSD of the central monomer g_1 as a function of time for a polymer with linear initial configuration simulated with the ssBFM (red). The MSD is rescaled using the centre of mass diffusion D_{cm}, which is obtained by fitting $g_3(t) = 6D_{cm}t$. To validate the simulations the results from BD simulations of the Rouse model are shown in blue. (**b**) The corresponding MSD of the centre of mass g_3.

confirming that the ssBFM both shows the properties of a FJC and successfully reproduces Rouse dynamics even when starting from a linear chain configuration.

Initialising the chain such that all segments are at their maximum extension we obviously expect deviations from the Rouse model, since the relaxation of the chain would be considerably slowed down again. This is confirmed by our simulations, as shown in fig. B.18. Since the configuration space is larger for the ssBFM with a spring potential, this model is expected to be more robust to different initial chain configurations. As shown in fig. B.17 however there is still a major slow down of the chain relaxation compared to the Rouse model. Nevertheless, as we do not intentionally initialise the polymer in a stretched out configuration during our target search simulations, the phantom chain implementation of the ssBFM yields an adequate compromise between fast and robust polymer dynamics. Also a linear configuration is not used to initialise the polymer chain in the target search simulations, but tethering of chain segments to the surface due to non-specific binding might lead to linear configurations of chain segments in the course of the simulation.

	$\langle \boldsymbol{R'} \rangle$	$\langle (\boldsymbol{R'} - \langle \boldsymbol{R'} \rangle)^2 \rangle$	$\langle \boldsymbol{R}_{ee}^2 \rangle$	$\langle \cos \rangle$	$\langle (\cos - \langle \cos \rangle)^2 \rangle$
	1.4092	0.0711	18.6498	−0.0046	0.3341
exact	1.4160	0	18.6923	0	0.3300

Table B.4: Comparing the ssBFM to the FJC model.

B5 Numerical Correction of the Polymer Diffusion Constant

It is well known [157, 167], that excluded volume interactions, a limited move set or segment length restrictions in a polymer model lead to a chain length dependence of the bead diffusion constant D_0. For the ssBFM of a phantom chain this leads to a dependence as shown in fig. B.22. While this is usually circumvented by rescaling the simulation results with the desired diffusion constant of a real polymer, this is not possible when the polymer is not the only moving part in the system. We therefore compute the monomer diffusion constant of a free polymer of varying length by fitting the centre of mass displacement $g_3(t) = 6D_0t/N$ and include a numerical correction in our target search simulations to ensure a constant rate D_0.

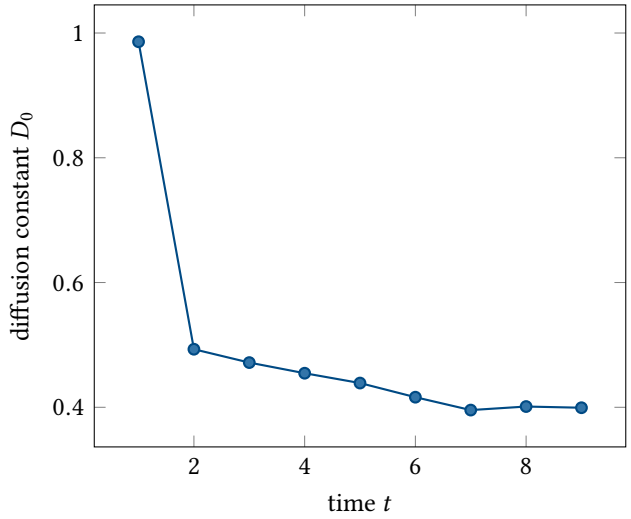


Figure B.22: Chain length dependent monomer diffusion. The monomer diffusion constant D_0 is shown as a function of chain length for the ssBFM of a phantom chain. It is obtained by fitting the centre of mass displacement $g_3(t) = 6D_0t/N$.

C Stochastic Processes

C1 Sequential Reversible Process

We consider a sequential reversible two-step process

$$S_1 \underset{k_2^-}{\overset{k_1^+}{\rightleftharpoons}} S_2 \overset{k_2^+}{\to} S_3 \,.$$

First we consider the general case of N states $n = \{1, \ldots, N\}$. The master equation for the transition probability $p(n,t|n',t')$ from state n' at time t' to state n at time t reads

$$\frac{\partial}{\partial t} p(n,t|n',t') = k_{n-1}^+ p(n-1,t|n',t') + k_{n+1}^- p(n+1,t|n',t') - (k_n^+ + k_n^-) p(n,t|n',t') \,. \quad \text{(C.1)}$$

Since we are interested in the time τ to reach the final state S_N, we set $k_N^- = 0$, such that S_N becomes an absorbing state. Define $G(m,t)$, the probability that at time t state S_N is not yet reached when starting from state m at time $t = 0$:

$$G(m,t) := \mathcal{P}\{\tau \geq t | n(0) = m\} = \sum_{n=1}^{N} p(n,t|m,0) \,.$$

The distribution of first passage times is obtained from $G(m,t)$, since

$$1 - G(m,t) = \mathcal{P}\{\tau < t | n(0) = m\} \,.$$

$1 - G(1,t)$ is the cumulative distribution function $\Phi(\tau)$ (CDF) of the first passage time, therefore

$$\Psi(\tau) = -\frac{\partial}{\partial t} G(1,t)|_{t=\tau} \,, \quad \text{(C.2)}$$

is the first passage time distribution. We exploit that the transition rates do not depend on time to write

$$G(m,t) = \sum_{n=1}^{N} p(n,0|m,-t) \,.$$

Therefore $G(m,t)$ follows from the evolution of p in the coordinate of the initial state, described by the backward master Equation

$$\frac{\partial}{\partial t'} p(n,t|n',t') = k_{n'}^+ (p(n,t|n',t') - p(n,t|n'+1,t')) + k_{n'}^- (p(n,t|n',t') - p(n,t|n'-1,t')) \,,$$

which leads to

$$-\frac{\partial}{\partial t}G(m,t) = k_m^+(G(m,t) - G(m+1,t)) + k_m^-(G(m,t) - G(m-1,t))\,, \tag{C.3}$$

with initial condition $G(m,0) = 1$ for $m < N$ and $G(N,0) = 0$.

We now go back to the two-step process $N = 3$. Since $G(3,t) = 0$ we have two equations:

$$\begin{aligned}
-\frac{\partial}{\partial t}G(1,t) &= k_1^+(G(1,t) - G(2,t)) + k_1^- G(1,t) \\
&= k_1^+(G(1,t) - G(2,t)) \\
-\frac{\partial}{\partial t}G(2,t) &= k_2^+(G(1,t) - G(3,t)) + k_2^-(G(2,t) - G(1,t)) \\
&= k_2^+ G(1,t) + k_2^-(G(2,t) - G(1,t))\,.
\end{aligned}$$

To solve this initial value problem, we use the Laplace transform

$$\begin{aligned}
F(s) &= \int_0^\infty f(t)e^{-st}\mathrm{d}t \\
f(t) &= \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} e^{st}F(s)\mathrm{d}s\,.
\end{aligned}$$

Using $\mathcal{L}(f') = sF(s) - f(0)$ we obtain

$$\begin{aligned}
-sg(1,s) + G(1,0) &= k_1^+(g(1,s) - g(2,s)) \\
-sg(1,s) + 1 &= k_1^+(g(1,s) - g(2,s)) \\
-sg(2,s) + G(2,0) &= k_2^+ g(2,s) + k_2^-(g(2,s) - g(1,s)) \\
-sg(2,s) + 1 &= k_2^+ g(2,s) + k_2^-(g(2,s) - g(1,s))\,.
\end{aligned}$$

We solve for $g(1,s)$ and $g(2,s)$ and obtain

$$g(1,s) = \frac{k_2^- + k_1^+ + k_2^+ + s}{k_2^- s + (k_1^+ + s)(k_2^+ + s)}\,,$$

as well as

$$g(2,s) = \frac{k_2^- + k_1^+ + s}{k_2^- s + (k_1^+ + s)(k_2^+ + s)}\,,$$

and after back transformation

$$G(1,t) = \frac{e^{-(a+b)t}(-2b + e^{2at}2b + 2a + e^{2at}2a)}{4a}\,, \tag{C.4}$$

and

$$G(2,t) = \frac{e^{-(a+b)t}(-2(b-k_2^+) + e^{2at}2(b-k_2^+) + 2a + e^{2at}2a)}{4a}\,, \tag{C.5}$$

with $a = \frac{1}{2}\sqrt{(k_2^- + k_1^+ + k_2^+)^2 - 4k_1^+k_2^+}$ and $b = \frac{1}{2}(k_2^- + k_1^+ + k_2^+)$. Using eq. (C.2) we obtain

$$\begin{aligned}\Psi(\tau) &= \frac{k_1^+k_2^+}{2a}\left(e^{(a-b)\tau} - e^{-(a+b)\tau}\right) \\ &= \frac{b^2 - a^2}{2a}\left(e^{(a-b)\tau} - e^{-(a+b)\tau}\right),\end{aligned}$$

and for the CDF

$$\Phi(\tau) = (1 - G(1,t))|_{t=\tau} = 1 - \frac{(a-b)e^{-(a+b)\tau} + (a+b)e^{(a-b)\tau}}{2a}.$$

Defining $\alpha = \frac{1}{a+b}$ and $\beta = \frac{1}{b-a}$, the CDF reads

$$\Phi(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}}}{\alpha - \beta}, \tag{C.6}$$

and the probability density function

$$\Psi(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}}{\alpha - \beta}. \tag{C.7}$$

When we assume that $k_2^- = 0$, it reduces to

$$\Phi(\tau) = 1 + \frac{k_2^+e^{-k_1^+\tau} - k_1^+e^{-k_2^+\tau}}{k_1^+ - k_2^+}. \tag{C.8}$$

When we assume that k_2^+ is rate limiting, it reduces to

$$\Phi(\tau) = 1 - e^{-k_2^+\tau}. \tag{C.9}$$

Mixed initial condition

The above first passage time distribution is based on a system that starts initially in state S_1. When we allow a mixed initial state, with a fraction x in state S_2 and a fraction $1 - x$ in state S_1, we obtain

$$\begin{aligned}\Phi(\tau) &= \mathcal{P}\{\tau < t | n(0) = xs_2 + (1-x)s_1\} \\ &= x\mathcal{P}\{\tau < t | n(0) = s_2\} + (1-x)\mathcal{P}\{\tau < t | n(0) = s_1\} \\ &= (1 - xG(2,t) - (1-x)G(1,t))|_{t=\tau},\end{aligned}$$

and

$$\Psi(\tau) = -\frac{\partial}{\partial t}(xG(2,t) + (1-x)G(1,t))|_{t=\tau}.$$

With $G(1,t)$ and $G(2,t)$ given in eqs. (C.4) and (C.5) this yields

$$\Phi(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}} - \alpha\beta c\left(e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}\right)}{\alpha - \beta}, \tag{C.10}$$

and

$$\Psi(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}} + c\left(\alpha e^{-\frac{\tau}{\beta}} - \beta e^{-\frac{\tau}{\alpha}}\right)}{\alpha - \beta},$$

with $c = xk_2^+$ and

$$\alpha = 2\left(k_1^+ + k_2^+ + k_2^- - \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+k_2^+}\right)^{-1},$$
$$\beta = 2\left(k_2^- + k_1^+ + k_2^+ + \sqrt{(k_2^- + k_1^+ + k_2^+)^2 - 4k_1^+k_2^+}\right)^{-1}.$$

Moments

For the sequential reversible process with fixed initial condition we use the moments of the probability density function eq. (C.7) to compute the mean first passage time and the variance as functions of α and β. Since α and β are not independent, we propagate the uncertainty with the whole covariance matrix $\sigma_{ij} = \mathrm{cov}[X_i, X_j] = \mathcal{E}[(X_i - \mathcal{E}[X_i])(X_j - \mathcal{E}[X_j]))]$, such that the uncertainty of a function $y(\alpha, \beta)$ reads

$$\begin{aligned}\Delta y &= \sqrt{\sum_{i=1}^{m}\left(\frac{\partial y}{\partial \sigma_{ii}}\right)^2 + 2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\partial y}{\partial \sigma_{ii}}\frac{\partial y}{\partial x_j}\sigma_{ij}} \\ &= \sqrt{\left(\frac{\partial y}{\partial \alpha}\sigma_\alpha\right)^2 + \left(\frac{\partial u}{\partial \beta}\sigma_\beta\right)^2 + 2\frac{\partial y}{\partial \alpha}\frac{\partial y}{\partial \beta}\sigma_{\alpha\beta}}.\end{aligned}$$

We compute the mean (first moment $\mu = \mathcal{E}[X]$)

$$\mu = \alpha + \beta$$
$$\Delta\mu = \sqrt{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}},$$

and the variance (second central moment $\sigma^2 = \mathcal{E}[(X - \mu)^2]$)

$$\sigma^2 = \alpha^2 + \beta^2$$
$$\Delta\sigma^2 = 2\sqrt{\alpha^2\sigma_\alpha^2 + \beta^2\sigma_\beta^2 + \alpha\beta\sigma_{\alpha\beta}}.$$

For the sequential reversible model with mixed initial condition, we proceed as above using the probability density function in eq. (C.10). The uncertainty of a function $y(\alpha, \beta, c)$ now reads

$$\Delta y = \sqrt{\left(\frac{\partial y}{\partial \alpha}\sigma_\alpha\right)^2 + \left(\frac{\partial u}{\partial \beta}\sigma_\beta\right)^2 + \left(\frac{\partial y}{\partial c}\sigma_c\right)^2 + 2\frac{\partial y}{\partial \alpha}\frac{\partial y}{\partial \beta}\sigma_{\alpha\beta} + 2\frac{\partial y}{\partial \alpha}\frac{\partial y}{\partial c}\sigma_{\alpha c} + 2\frac{\partial y}{\partial \beta}\frac{\partial y}{\partial c}\sigma_{\beta c}}.$$

We again compute the mean

$$\begin{aligned}\mu &= \alpha + \beta - \alpha\beta c \\ \Delta\mu &= \sqrt{\sigma_\alpha^2(1-\beta c)^2 + \sigma_\beta^2(1-\alpha c)^2 + \sigma_c^2\alpha^2\beta^2 + 2\sigma_{\alpha\beta}(1-\alpha c)(1-\beta c) - 2\alpha\beta(\sigma_{\alpha c}(1-\beta c) + \sigma_{\beta c}(1-\alpha c))}\,,\end{aligned}$$

and the variance

$$\begin{aligned}\sigma^2 =& \alpha^2 + \beta^2 - \alpha^2\beta^2c^2 \\ \Delta\sigma^2 =& \Big\{4\sigma_\alpha^2\alpha^2(\beta^2c^2-1)^2 + 4\sigma_\beta^2\beta^2(\alpha^2c^2-1)^2 + \sigma_c^2(\beta^2+\alpha^2c(1-\beta^2c^2))^2 + 8\sigma_{\alpha\beta}\alpha\beta(\alpha^2c^2-1)(\beta^2c^2-1) \\ &+8\sigma_{\alpha c}\alpha^3\beta^2c(\beta^2c^2-1) + 8\sigma_{\beta c}\alpha^2\beta^3c(\alpha^2c^2-1)\Big\}^{\frac{1}{2}}\,.\end{aligned}$$

C2 Simultaneous Two Step Process

We consider a two channel process

$$S_1 \underset{k_2^-}{\overset{k_1^+}{\rightleftharpoons}} S_2 \underset{k_3^-}{\overset{k_2^+}{\rightleftharpoons}} S_3\,,$$

Where the end state S_3 can be reached via two different pathways. For the general case of N states $n = \{1,\dots,N\}$ we therefore have the same master equation as shown in eq. (C.1). With the mixed initial condition

$$\Phi(\tau) = (1 - xG(3,t) - (1-x)G(1,t))|_{t=\tau}\,,$$

and

$$\Psi(\tau) = -\frac{\partial}{\partial t}(xG(3,t) + (1-x)G(1,t))|_{t=\tau}\,.$$

with eq. (C.3) and initial condition $G(1,0) = G(3,0) = 1$ and $G(2,0) = 0$, $k_1^- = k_2^+ = 0$ (absorbing state) and $G(2,t) = 0$ (absorbing state) we obtain

$$\begin{aligned}-\frac{\partial}{\partial t}G(1,t) &= k_1^+(G(1,t) - G(2,t)) + k_1^-(G(1,t) - G(0,t)) \\ &= k_1^+G(1,t) \\ -\frac{\partial}{\partial t}G(2,t) &= k_2^+(G(2,t) - G(3,t)) + k_2^-(G(2,t) - G(1,t)) \\ &= 0 \\ -\frac{\partial}{\partial t}G(3,t) &= k_3^+(G(3,t) - G(4,t)) + k_3^-(G(3,t) - G(2,t)) \\ &= k_3^-G(3,t)\,.\end{aligned}$$

To solve this initial value problem, we use the Laplace transform. Using $\mathcal{L}(f') = sF(s) - f(0)$ we obtain

$$-sg(1,s) + G(1,0) = k_1^+g(1,s)$$

$$
\begin{aligned}
-sg(1,s) + 1 &= k_1^+ g(1,s) \\
g(1,s) &= \frac{1}{s + k_1^+} \\
-sg(3,s) + 1 &= k_3^- g(3,s) \\
g(3,s) &= \frac{1}{s + k_3^-} .
\end{aligned}
$$

And after back transform

$$
\begin{aligned}
G(1,t) &= e^{-k_1^+ t} \\
G(3,t) &= e^{-k_3^- t} .
\end{aligned}
$$

And the probability density function is

$$
\Psi(\tau) = (x-1)k_1^+ e^{-k_1^+ \tau} + x k_3^- e^{-k_3^- \tau} ,
$$

with the cumulative distribution

$$
\begin{aligned}
\Phi(\tau) &= (1 - xG(3,t) - (1-x)G(1,t))|_{t=\tau} \\
&= 1 - xe^{-k_3^- \tau} + (x-1)e^{-k_1^+ \tau} ,
\end{aligned}
$$

and the mean

$$
\begin{aligned}
\mu &= \frac{k_3^- - x k_3^- + x k_1^+}{k_3^- k_1^+} \\
&= \frac{x}{k_3^-} + \frac{1-x}{k_1^+} .
\end{aligned}
$$

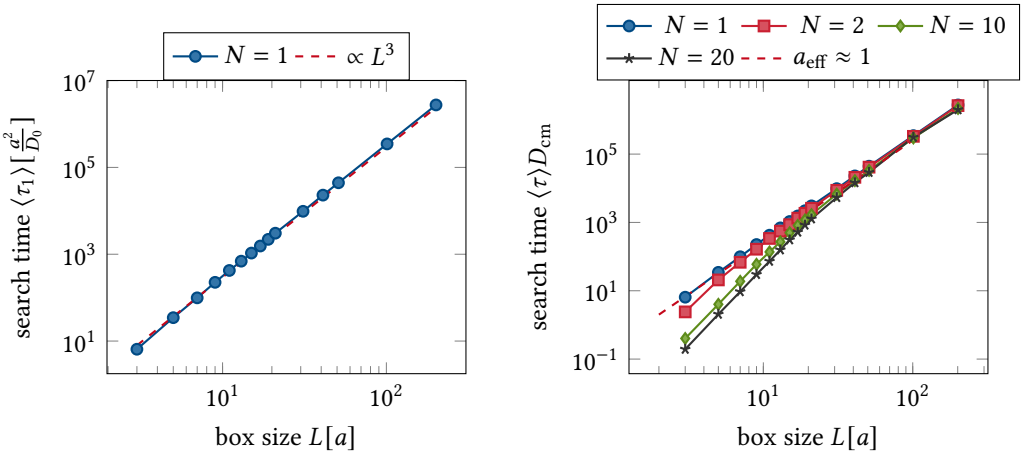
Search Dynamics of a Polymer in Confinement

D

We show additional plots and tables investigating the target search process of surface proteins for a binding site on a polymer inside the volume by kinetic Monte Carlo simulations.

N	a_{eff}	N	a_{eff}	N	a_{eff}
1	0.738 956 ± 0.000 010	6	0.8652 ± 0.0019	15	1.1093 ± 0.0062
2	0.772 22 ± 0.000 28	7	0.829 80 ± 0.000 39	20	1.0137 ± 0.0077
3	0.6642 ± 0.0018	8	0.7670 ± 0.0028	25	1.1364 ± 0.0043
4	0.710 99 ± 0.000 88	9	0.8362 ± 0.0014	30	0.9820 ± 0.0018
5	0.6796 ± 0.0029	10	0.9368 ± 0.0025	35	1.397 ± 0.012

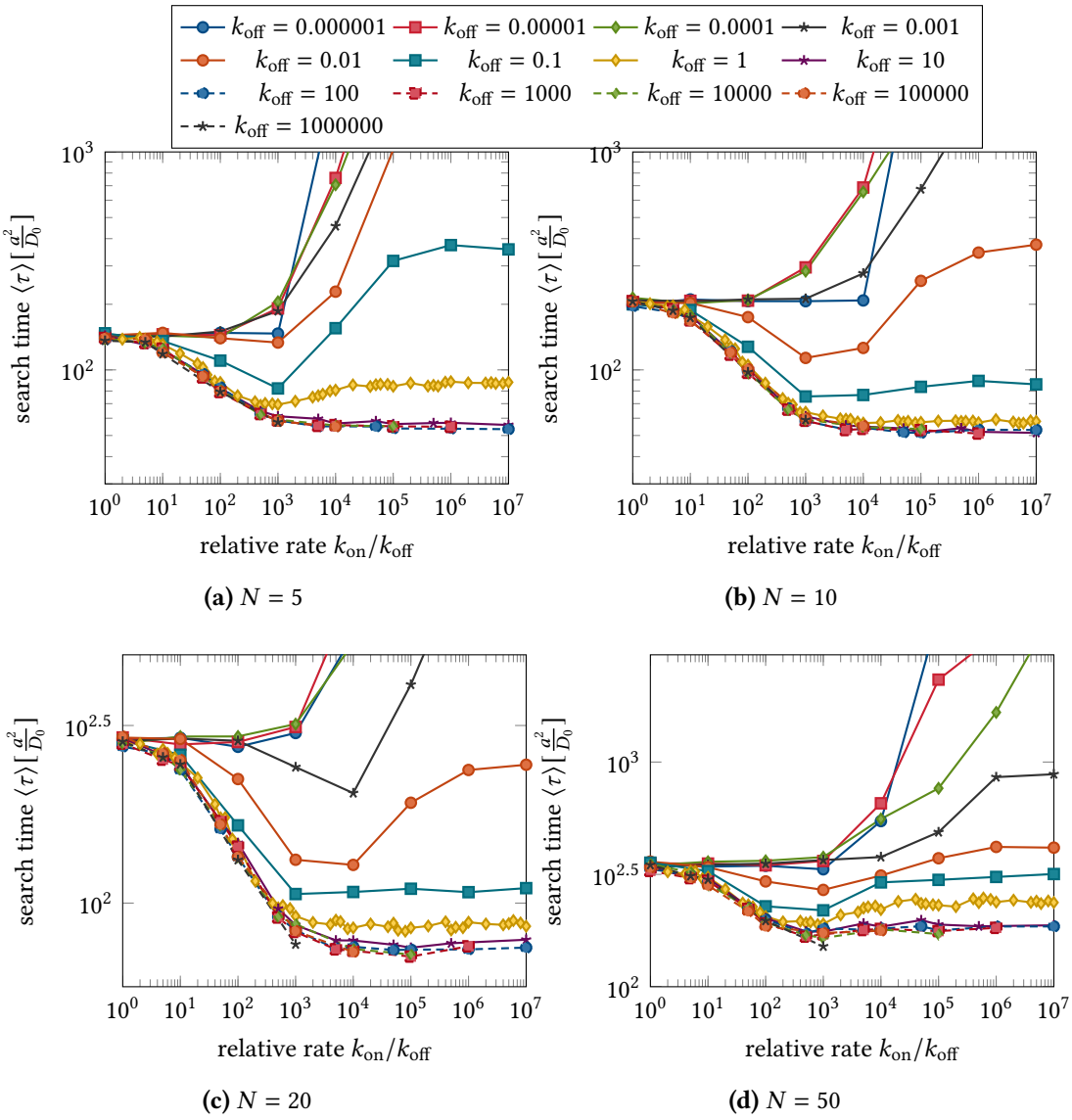
Table D.1: Fit results. Results from fitting the search time normalised to the polymer length to $\langle\tau\rangle/N = V/(4D_0 a_{\text{eff}})$ in order to obtain the effective target size a_{eff}. Fits were obtained using the *curve_fit* function of the scipy module in Python, choosing a Levenberg-Marquardt method.



(a) Monomer search time

(b) Polymer search time

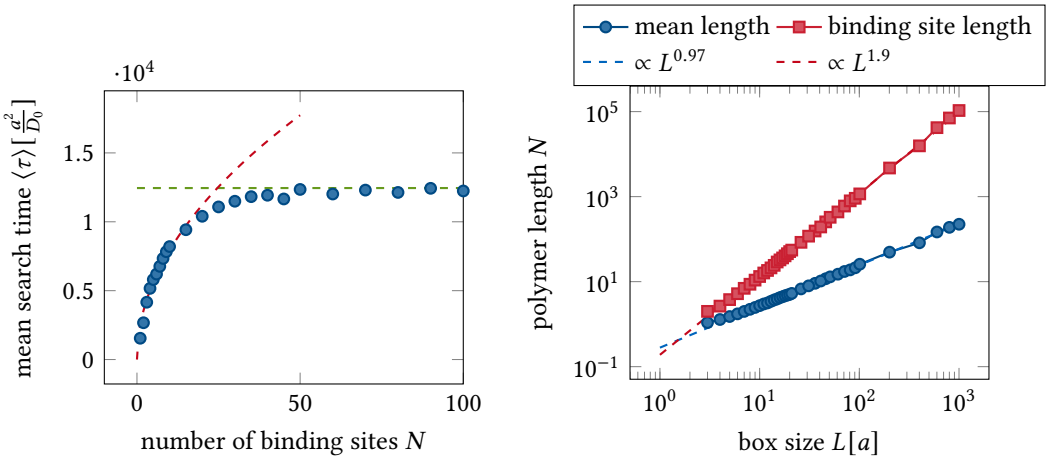
Figure D.1: Search time of monomers and polymers. (**a**) Mean search time $\langle\tau\rangle$ of a polymer of length $N = 1$ for varying the box size L. The search time of a monomer, i.e. the narrow escape time (NET) of a Brownian particle is given by $\langle\tau\rangle \approx V/(4D_0 a_t)$, with the volume V, the diffusion constant D_0 and the target size a_t. (**b**) Mean search time $\langle\tau\rangle$ of a polymer of length N for varying the box size L. The search time of a small polymer (compared to the confining volume) is given by $\langle\tau\rangle \approx V/(4D_0 a_{\text{eff}})$, with the volume V, the diffusion constant D_0 and the effective target size a_{eff}.



(a) $N = 5$ **(b)** $N = 10$

(c) $N = 20$ **(d)** $N = 50$

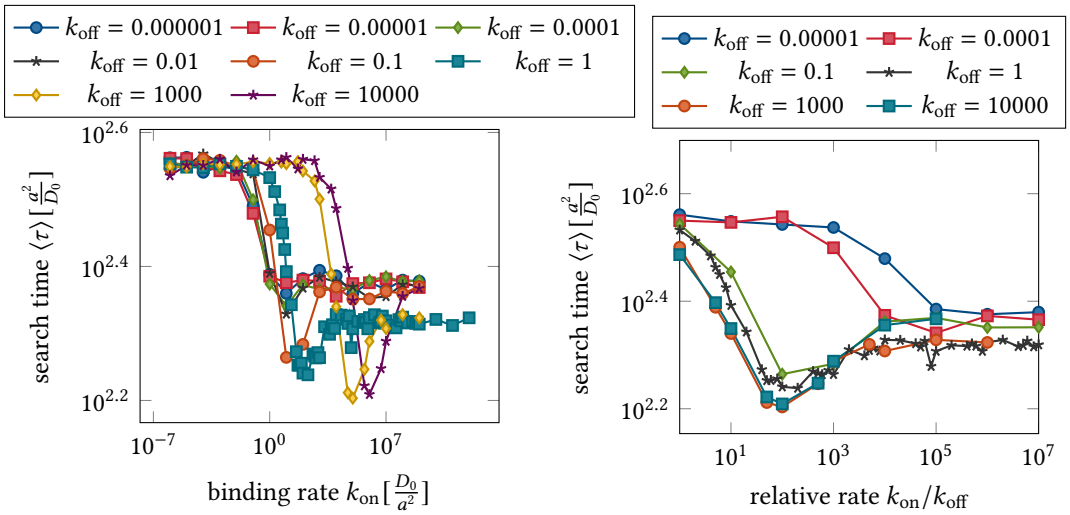
Figure D.2: Polymer length dependence of non-specific binding. Mean search time $\langle\tau\rangle$ as a function of the relative binding rate k_{on}/k_{off} for varying unbinding rate and different lengths of the polymer N. Increasing the polymer length has a similar effect as decreasing the size of the simulation box, since the behaviour essentially depends on the polymer density in the volume.



(a) Fitting the crossover chain length

(b) Average segment length

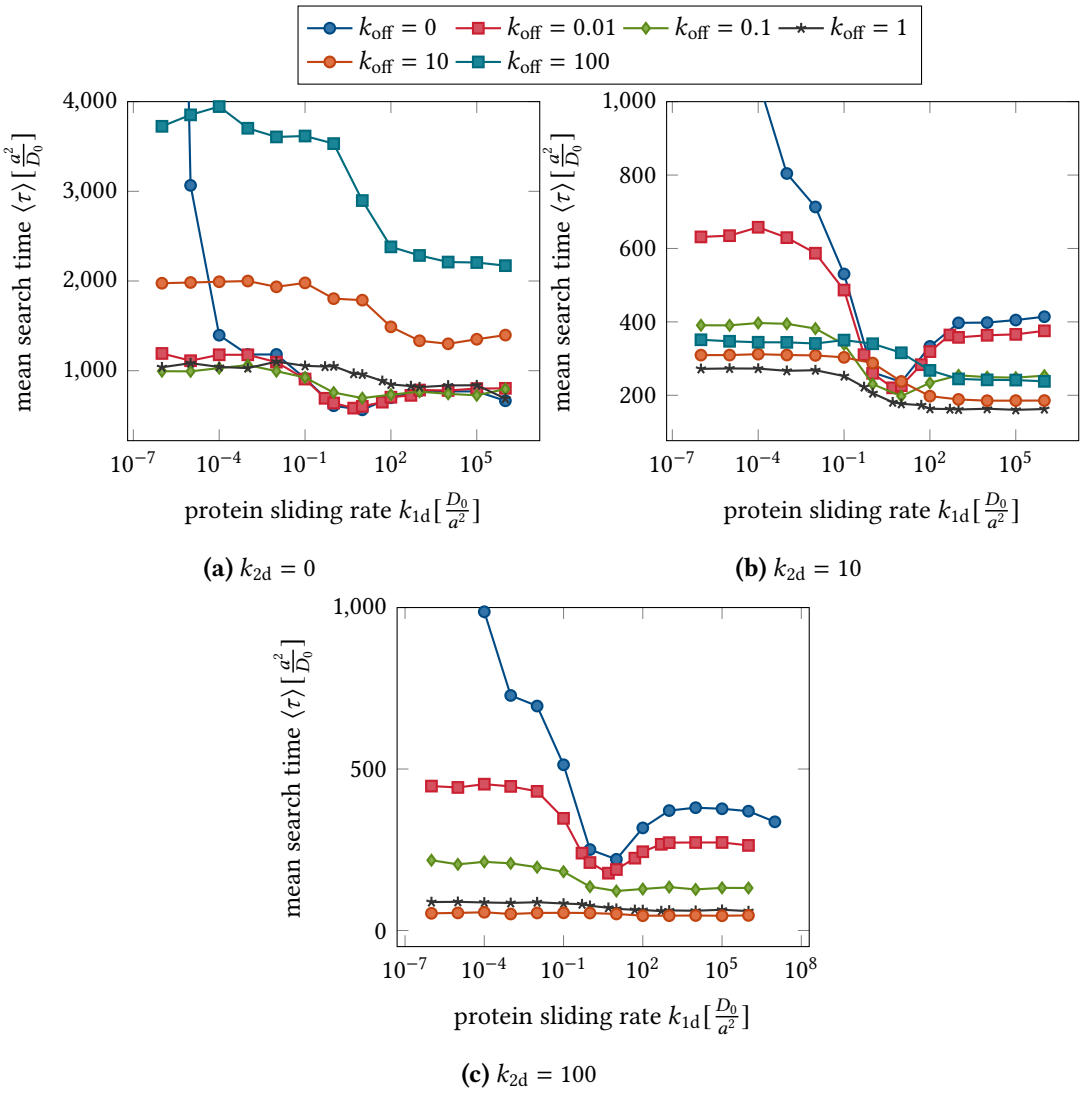
Figure D.3: Fitting the crossover length and average segment length. (**a**) Fitting the initial increase (the range $N = 1$ to L) of the mean search time $\langle\tau\rangle$ to $f(N) = c\sqrt{N}$ (red dashed line) and taking the intersect with the fit to the plateau value (green dashed line) as the crossover polymer length. (**b**) The blue data show the average segment size between two encounters with the boundary. The red data show the average segment size for a few binding sites along the polymer if each encounter with the boundary would cut the polymer. The dashed lines correspond to the best fit to the data.



(a) Binding rate

(b) Relative binding rate

Figure D.4: Non-specific binding and unbinding with sliding. (**a**) Mean search time $\langle\tau\rangle$ as a function of the binding rate k_{on} for different values of k_{off} and constant sliding rate $k_{1d} = 10$. Sliding has the effect that binding is always favourable, such that the search time plateau for small binding rates is higher than the plateau for large binding rates. (**b**) Mean search time as a function of the relative binding rate k_{on}/k_{off}. The minimum at $k_{on}/k_{off} \approx 100$ is unchanged.



(a) $k_{2d} = 0$

(b) $k_{2d} = 10$

(c) $k_{2d} = 100$

Figure D.5: Protein rate dependence of sliding. Mean search time $\langle\tau\rangle$ as a function of the sliding rate k_{1d} for varying unbinding rate and different values of the 2d diffusion rate of the protein k_{2d}. The position of the minimum is not affected by the protein diffusion rate and also the factor between the plateaus is conserved. Only the absolute height of the plateaus is affected due to the speedup of the process by faster protein diffusion.

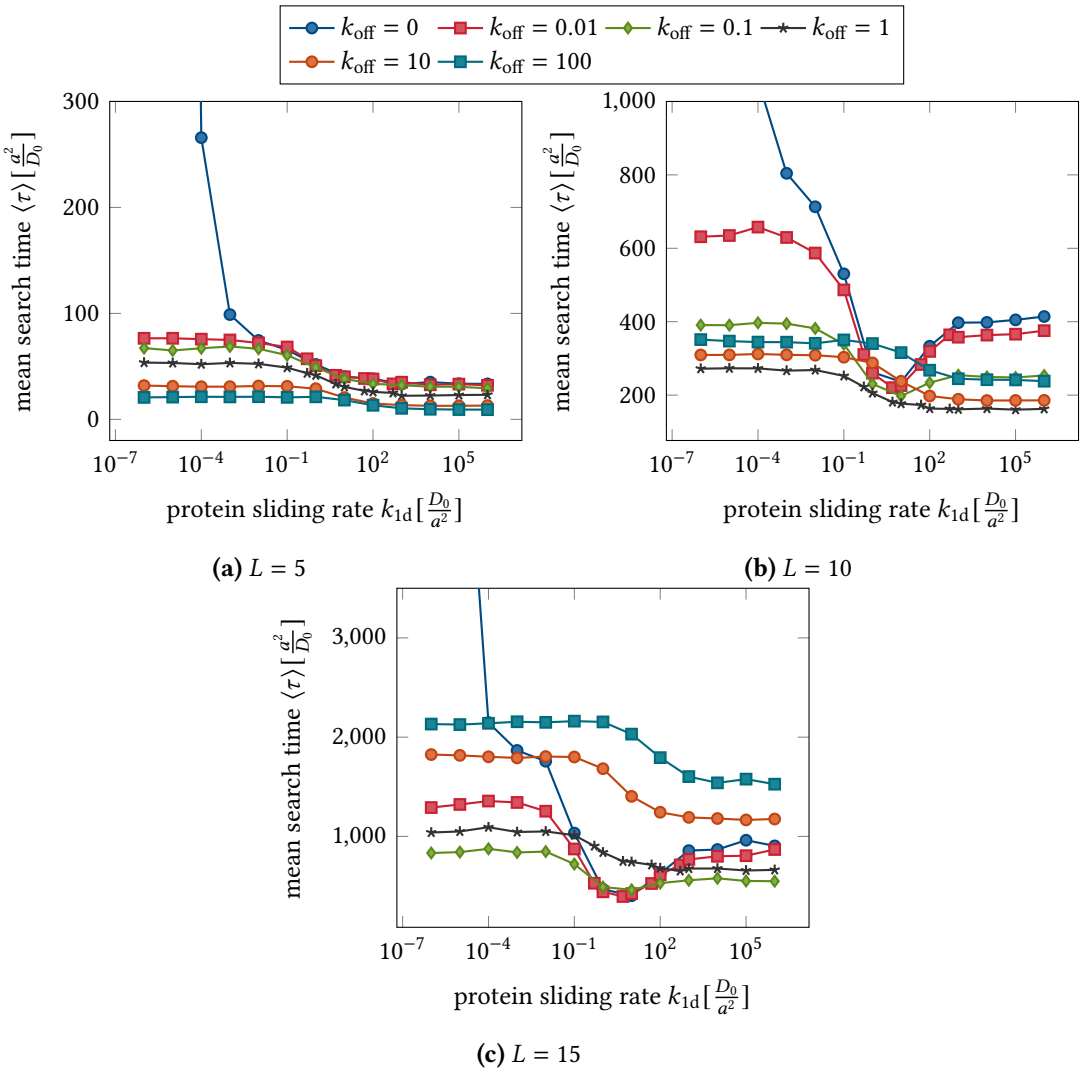
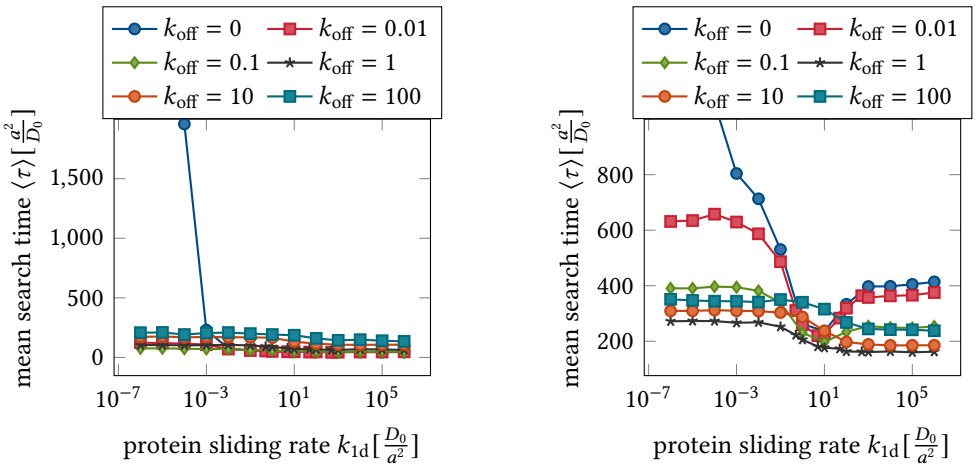
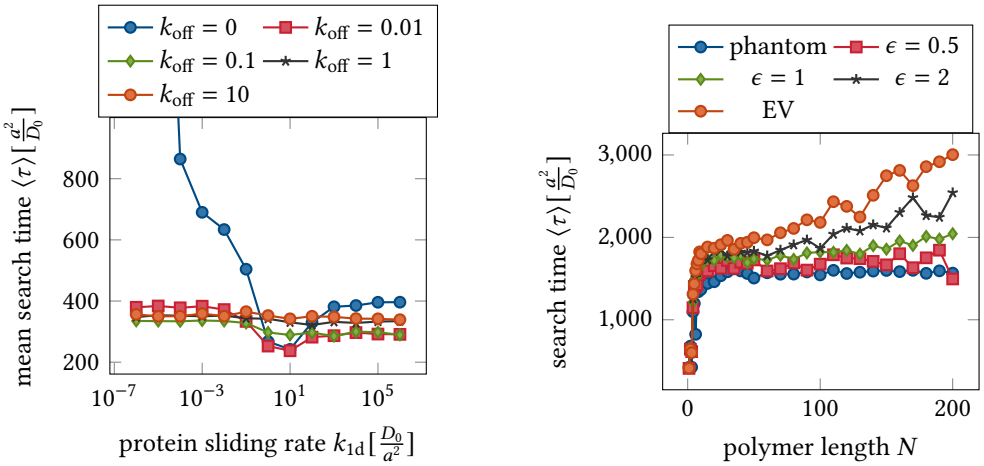


Figure D.6: Volume size dependence of sliding. Mean search time $\langle\tau\rangle$ as a function of the sliding rate k_{1d} for varying unbinding rate and different volume sizes of the cubic grid. The volume size does in general not affect the position of the minimum or the factor between the two plateaus. Only for very small boxes the minimum disappears and the factor between the plateaus is larger, because at edges and corners the probability for a neighbouring bead to stay at the surface is larger than at a face centre, leading to a larger fraction of polymer beads at the surface, along which the bound protein can slide. The minimum for approximately permanent binding disappears for very small boxes probably because a large fraction of the polymer is at the surface anyways and therefore it does not slow down polymer diffusion if it is tethered to the surface by the sliding protein.



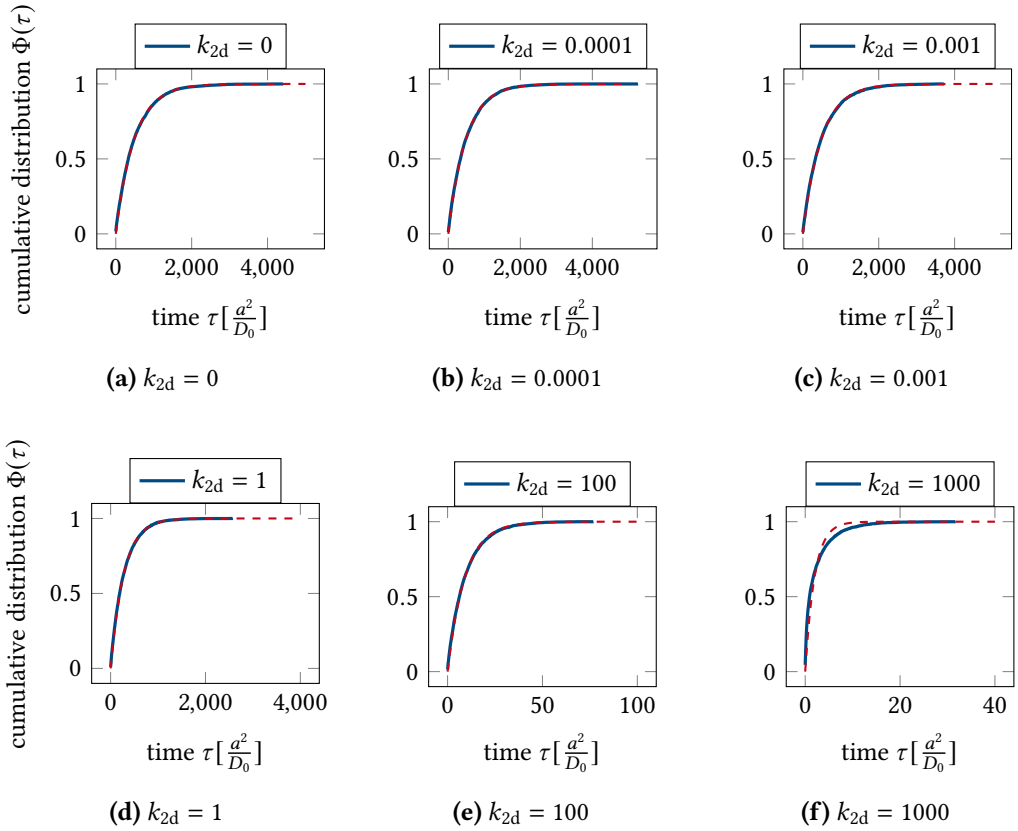
(a) $N = 10$ **(b)** $N = 100$

Figure D.7: Polymer length dependence of sliding. Mean search time $\langle\tau\rangle$ as a function of the sliding rate k_{1d} for varying unbinding rate and different lengths of the polymer N. Varying the polymer length does not affect the position of the minimum or the factor between the plateaus. Only for very short polymers the minimum for approximately permanent binding disappears, likely because the positive effect of bringing the binding site close to the surface due to tethering dominates over the slowdown of polymer diffusion.



(a) Effect of sliding **(b)** Effect of self-exclusion

Figure D.8: Effect of sliding and self-exclusion. **(a)** The mean search time $\langle\tau\rangle$ is shown as a function of the protein sliding rate k_{1d} for different values of the unbinding rate and constant binding rate $k_{on} = 1$. Choosing a different binding rate does not affect the position of the minimum or the factor between the plateaus. **(b)** Search time $\langle\tau\rangle$ for varying polymer length N and different values of the energy penalty ϵ, ranging from $\epsilon = 0$ (phantom chain) to $\epsilon = \infty$ (self-avoiding chain). The confining volume is a cube of size $L = 10$.



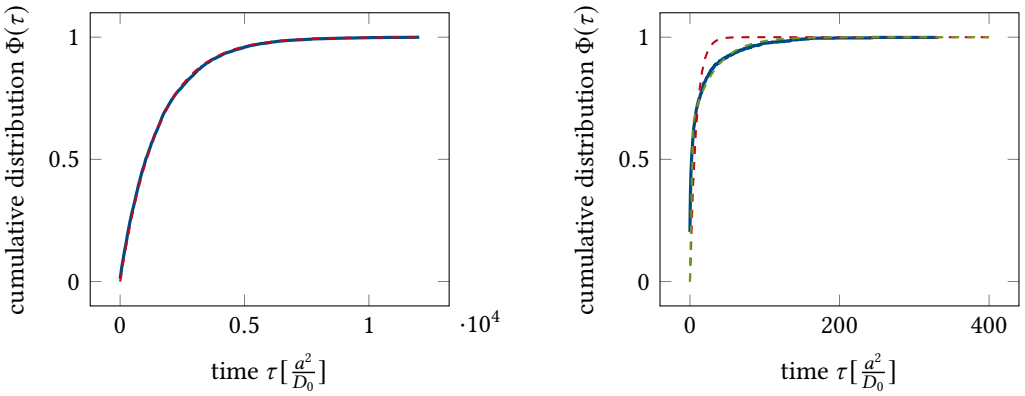
(a) $k_{2d} = 0$ **(b)** $k_{2d} = 0.0001$ **(c)** $k_{2d} = 0.001$

(d) $k_{2d} = 1$ **(e)** $k_{2d} = 100$ **(f)** $k_{2d} = 1000$

Figure D.9: CDF for a monomer. (**a**)-(**f**) Cumulative distribution $\Phi(\tau)$ of the first passage time for varying mobility of the protein and a polymer of length $N = 1$. The red dashed lines show the best fit to an exponential distribution. We observe two regimes: while for a very fast protein the CDF deviates from an exponential distribution, they agree very well for $k_{2d} \leq 100$. As expected, since for a monomer the corner effect is almost negligible, there is no deviation from the single exponential distribution for an immobile or very slowly diffusing protein.

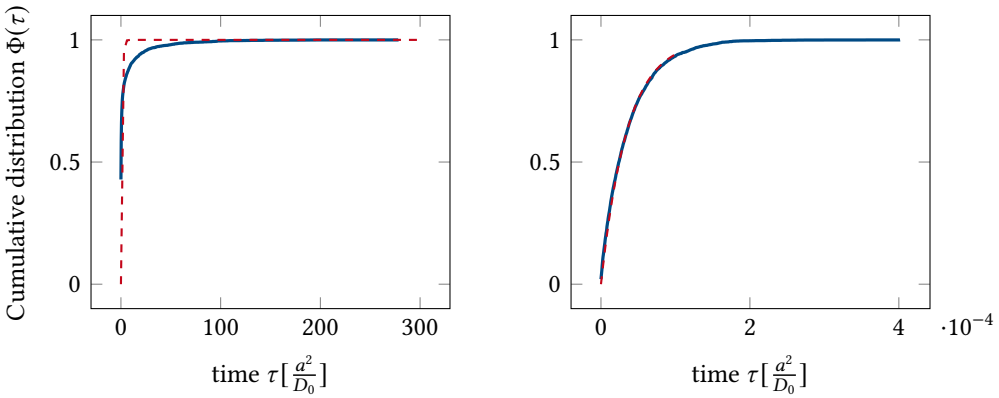
exponent	**fitted value**	**literature value**
$\alpha_{0,c}$	1.161 ± 0.014	1.45
$\alpha_{0,c}^{mid}$	1.959 ± 0.016	
$\alpha_{0,f}$	0.493 ± 0.012	0.5
$\alpha_{0,f}^{mid}$	0.7931 ± 0.0082	1.01

Table D.2: Fit results. Results from fitting the mean search time of an end monomer or a mid monomer respectively to reach the corner or the face centre of the cubic box to $\langle\tau\rangle(N) = N^{\alpha_i}$. Fits were obtained using the *curve_fit* function of the scipy module in Python, choosing a Levenberg-Marquardt method.



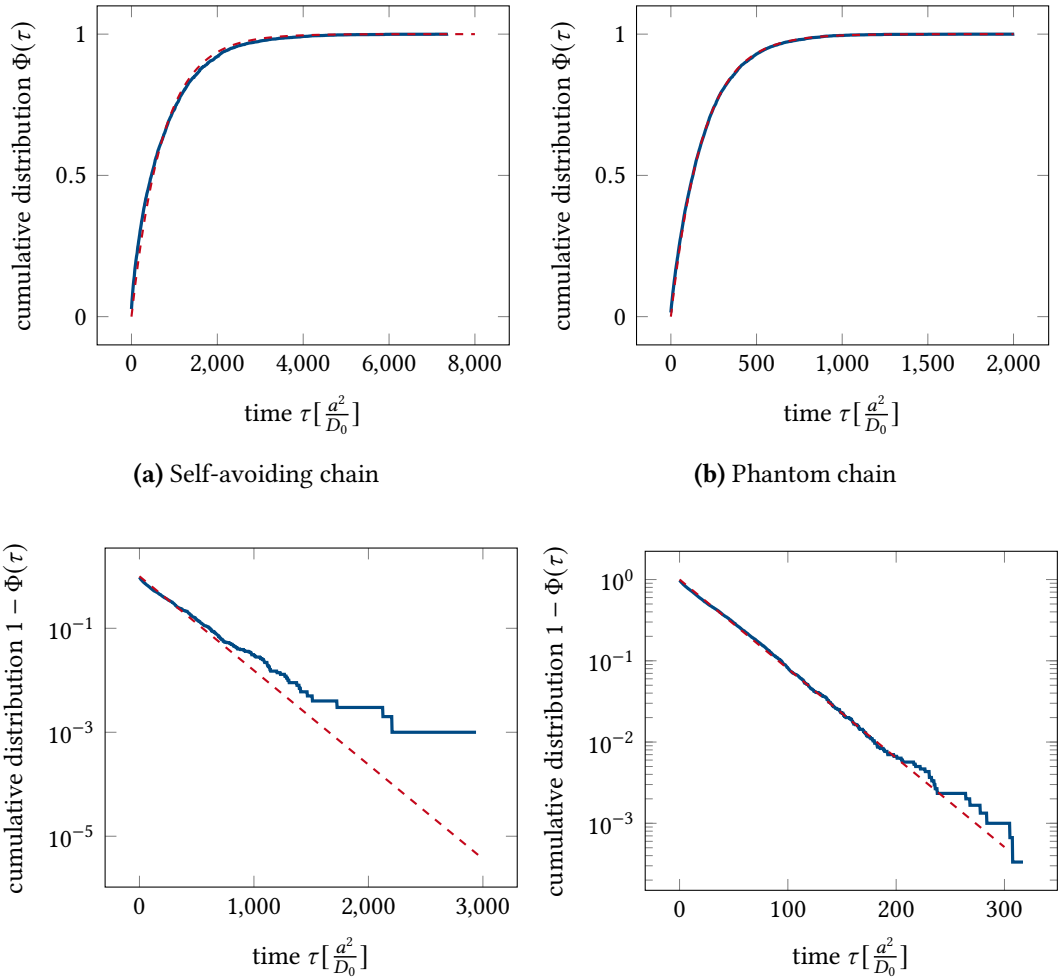
(a) Static protein **(b)** Sequential process

Figure D.10: CDF of a static protein and fit to a sequential process. (**a**) Cumulative distribution $\Phi(\tau)$ of the first passage time for a static protein placed at the face centre of the cubic box. The red dashed line shows the best fit to a single exponential distribution. As expected, the search process for a static protein can be well described by a single exponential (Poisson process). (**b**) Cumulative distribution $\Phi(\tau)$ of the first passage time for a moving protein ($k_{2d} = 1000$). The green dashed line shows the best fit to the CDF of a sequential two-step process with mixed initial condition, the red dashed line shows the best fit to a single exponential distribution. As expected, the agreement with the sequential two-step process is much better than the accordance with the single exponential distribution.



(a) $k_{2d} = 1000$ **(b)** $k_{2d} = 10000000$

Figure D.11: Binding site at the surface. Cumulative distribution $\Phi(\tau)$ of the first passage time for varying mobility of the protein. The red dashed lines show the best fit to an exponential distribution. (**a**) For intermediate protein diffusion rates a deviation from the exponential distribution arises because both the polymer and the protein contribute to the search. (**b**) As expected, when the binding site on the polymer is placed at the surface, the distribution agrees well with a single exponential for a very fast diffusing protein.



(a) Self-avoiding chain

(b) Phantom chain

(c) Self-avoiding chain

(d) Phantom chain

Figure D.12: Effect of self-exclusion on the distribution. The cumulative distribution function is shown for $L = 4$ and $N = 50$ with self-exclusion ((**a**), (**c**)) and without self-exclusion ((**b**), (**d**)). The distribution with self-exclusion shows the onset of heavy tails, because some of the conformations of the densely packed self-avoiding polymer allow for very few movement. Due to the massive slowdown of the simulations in this regime, higher polymer densities could not be simulated with excluded volume interactions.

Search Dynamics of a Membrane-Integrated One-Component Receptor

E

E1 Construction of Strains and Plasmids

Molecular methods were carried out according to standard protocols or according to the manufacturer's instructions. Kits for the isolation of plasmids and the purification of PCR products were purchased from Süd-Laborbedarf (SLG; Gauting, Germany). Enzymes were purchased from New England BioLabs (Frankfurt, Germany). Bacterial strains and plasmids used in this study are summarised in table E.1.

E. coli strains were cultivated in LB medium (10 g l^{-1} NaCl, 10 g l^{-1} tryptone, 5 g l^{-1} yeast extract) or in Kim Epstein (KE) medium [209] adjusted to pH 5.8 or pH 7.6, using the corresponding phosphate-buffer. *E. coli* strains were always incubated aerobically in a rotary shaker at 37 °C. KE medium was always supplemented with 0.2000 %(w/v) glucose. Generally, lysine was added to a final concentration of 10 mM unless otherwise stated. If necessary, media were supplemented with 100 µg ml^{-1} ampicillin or 50 µg ml^{-1} kanamycin sulfate. To allow the growth of the conjugation strain *E. coli* WM3064, we added meso-diamino-pimelic acid (DAP) to a final concentration of 200 µM.

In order to gain strain *E. coli* MG1655-*parS_ori*, the *parS* site of *Yersinia pestis* was inserted close to the *ori*, at 84.3′ in *E. coli* MG1655. Briefly, the *parS* region was inserted between *pstS* and *glmS*. Therefore, DNA fragments comprising 650 bp of *pstS* and *glmS* and the *parS* region were amplified by PCR using MG1655 genomic DNA as template and the plasmid pFH3228, respectively. After purification, these fragments were assembled via Gibson assembly [210] into EcoRV-digested pNPTS138-R6KT plasmid, resulting in the pNTPS138-R6KT-*parS_ori* plasmid. The resulting plasmid was introduced into *E. coli* MG1655 by conjugative mating using *E. coli* WM3064 as a donor on LB medium containing DAP. Single-crossover integration mutants were selected on LB plates containing kanamycin but lacking DAP. Single colonies were then streaked out on LB plates containing 10 %(w/v) sucrose but no NaCl to select for plasmid excision. Kanamycin-sensitive colonies were then checked for targeted insertion by colony PCR and sequencing of the respective PCR fragment. In order to gain strain *E. coli* MG1655Δ*cadC-parS_ori*, the *parS* site of *Y. pestis* was inserted close to the *ori*, at 84.3′ in *E. coli* MG1655Δ*cadC*, as described above.

In order to gain strain *E. coli* MG1655_P_{cadBA}_terminus, the *cadBA* promoter region was inserted at the terminus (33.7′) in *E. coli* MG1655. Construction of this strain was achieved via double homologous recombination using the pNTPS138-R6KT-P_{cadBA}_terminus plasmid [81] as described above. Correct colonies were then checked for targeted insertion by colony PCR and sequencing of the respective PCR fragment.

Details of the strains and plasmids used in this study are summarised in table E.1.

Strains	**Relevant genotype or description**	**Reference**
E. coli DH5αλpir	*endA1 hsdR17 glnV44 (= supE44) thi-1 recA1 gyrA96 relA1 ψ80'lacΔ* (*lacZ*)M15 Δ((*lacZYA-argF*)U169 *zdg*-232::Tn10 *uidA::pir*$^{+}$	[211]
E. coli WM3064	*thrB*1004 *pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)567 Δ*dapA*1341::[*erm pir*]	[212]
E. coli MG1655 (N-P_{cadBA})	K-12 F^{-} λ^{-} *ilvG*$^{-}$ *rfb*-50 *rph-1*	[192]
E. coli MG1655_ P_{cadBA}_terminus (N+T-P_{cadBA})	Additional *cadBA* promoter region at the terminus (33.7′) in MG1655	This work
E. coli MG1655 ΔP_{cadBA}_ P_{cadBA}_terminus (T-P_{cadBA})	Clean deletion of *cadBA* promoter region in MG1655 with relocated *cadBA* promoter region at the terminus (33.7′)	[81]
E. coli MG1655-*parS_ori*	Insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* (84.3′) in MG1655	This work
E. coli MG1655 Δ*cadC*	Deletion of *cadC* gene in MG1655, Km^{R}	[213]
E. coli MG1655 Δ*cadC-parS_ori*	Insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* (84.3′) in MG1655Δ*cadC*, Km^{R}	This work
Plasmids	**Relevant genotype or description**	**Reference**
pET-*mCherry-cadC*	N-terminal fusion of *cadC* with *mCherry*, connected with a 22 amino acid long linker containing a 10His tag in pET16b, Amp^{R}	[81]
pNTPS138-R6KT-P_{cadBA}_terminus	pNPTS-138-R6KT-derived suicide plasmid for insertion of *cadBA* promoter region at terminus in *E. coli* MG1655, Km^{R}	[81]
pNTPS138-R6KT-*parS_ori*	pNPTS-138-R6KT-derived suicide plasmid for insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* in *E. coli* MG1655, Km^{R}	This work
pFHC2973	N-terminal fusion of *parB* with *ygfp*, Amp^{R}	[214]
pFH3228	Plasmid carrying the pMT1-*parS* of *Yersinia pestis*, Amp^{R}	[214]

Table E.1: Strains and plasmids used in this study.

E2 *In Vivo* Fluorescence Microscopy

To analyse search response of mCherry-CadC to its binding site(s), overnight cultures of *E. coli* MG1655 (one CadC binding site close to *ori*, N-P_{cadBA}),

E. coli MG1655ΔP$_{cadBA}$_P$_{cadBA}$_terminus (one CadC binding site at terminus, T-P$_{cadBA}$) and *E. coli* MG1655_P$_{cadBA}$_terminus (two CadC binding sites, N+T-P$_{cadBA}$), each carrying pET-*mCherry-cadC*, were prepared in KE medium pH 7.6 and aerobically cultivated at 37 °C. The overnight cultures were used to inoculate day cultures (OD_{600} of 0.1) in fresh medium at pH 7.6. At an OD_{600} of 0.5, cells were gently centrifuged and resuspended, thereby exposing them to low pH (KE medium pH 5.8 + lysine). Then the cultures were aerobically cultivated at 37 °C and every 1 min after the shift to low pH, 2 µl of the culture was spotted on 1 %(w/v) agarose pads (prepared with KE medium pH 5.8 + lysine), placed onto microscope slides and covered with a coverslip. Subsequently, images were taken on a Leica DMi8 inverted microscope equipped with a Leica DFC365 FX camera (Wetzlar, Germany). An excitation wavelength of 546 nm and a 605 nm emission filter with a 75 nm bandwidth was used for mCherry fluorescence with an exposure of 500 ms, gain 5, and 100 % intensity. Before shifting the cells to low pH, 2 µl of the cultures in KE medium pH 7.6 were spotted on 1 %(w/v) agarose pads (prepared with KE medium pH 7.6) and imaged as a control.

To analyse the spatiotemporal localisation of a chromosomal locus, the *parS* site was inserted close to the *ori*. The localisation of the *parS* site was visualised via the binding of ParB-yGFP [214]. *E. coli* MG1655-*parS_ori* cells carrying plasmid pFH3228 were cultivated in KE medium pH 7.6 as described above. At an OD_{600} of 0.5, 2 µl of the culture were shifted on 1 %(w/v) agarose pads (prepared with KE medium pH 7.6 or pH 5.8 + lysine) and placed onto microscope slides and covered with a coverslip. Subsequently, every 30 s time-lapse images of the same cells were taken on a Leica DMi8 inverted microscope equipped with a Leica DFC365 FX camera (Wetzlar, Germany) of the same positions. An excitation wavelength of 485 nm and a 510 nm emission filter with a 75 nm bandwidth were used for ParB-yGFP fluorescence with an exposure of 350 ms, gain 3, and 100 % intensity.

E3 Fitting the Experimental Data

The results from fitting the experimentally computed CDF to the theoretical models are shown. Moreover, in fig. E.1 we show the data of the fraction of cells with clusters $\nu(t)$ for the three *E. coli* strains, from which the response function is computed.

Strain	α **[min]**	β **[min]**	c **[min^{-1}]**	$\langle\tau\rangle$ **[min]**	σ^2 **[min^2]**
N-P$_{cadBA}$	7.87 ± 0.60	0.52 ± 0.13	0.87 ± 0.15	4.84 ± 0.19	49.6 ± 8.8
T-P$_{cadBA}$	4.20 ± 0.26	6×10^{-15}±0.1800		4.20 ± 0.15	17.6 ± 2.0
N+T-P$_{cadBA}$	2.02 ± 0.16	1.1000×10^{-14}±0.2200		2.02 ± 0.12	4.09 ± 0.66

Table E.2: Fit results. Results from fitting the experimentally computed CDF to the sequential reversible model with mixed initial condition (N-P$_{cadBA}$) and fixed initial condition (T-P$_{cadBA}$ and N+T-P$_{cadBA}$). The fit parameters α, β and c were used to compute the mean first passage time and the variance with uncertainties obtained from error propagation using the full covariance matrix.

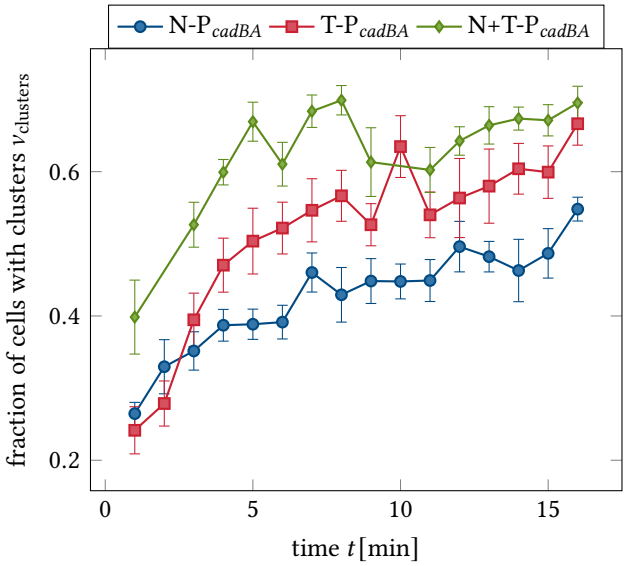


Figure E.1: Dynamics of the target search by CadC. Fluorescent microscopy images were taken every minute after receptor activation and analysed for CadC clusters for all three *E. coli* strains. The plot shows the fraction of cells with clusters $\nu(t)$ as a function of time t after the medium shift to low pH and lysine.

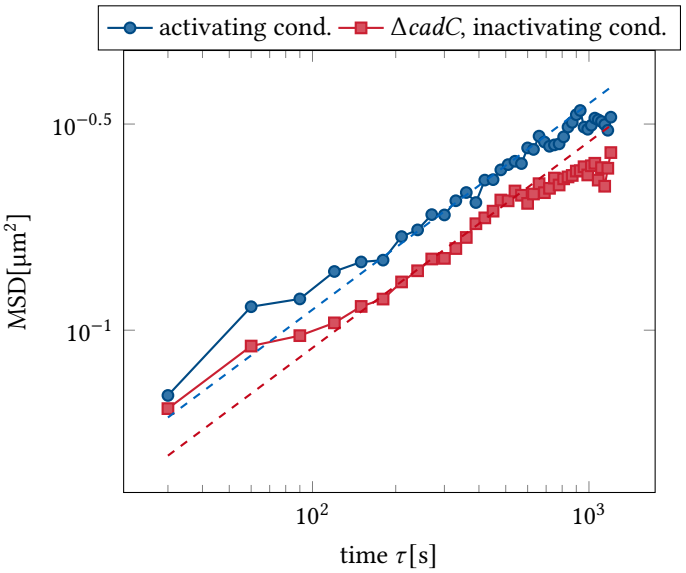


Figure E.2: Mean squared displacement of ParB foci. The MSD of ParB foci was calculated by selecting the closest foci in subsequent image frames and calculating the ensemble averaged mean squared displacement as a function of time lag τ. The dashed lines show the fit to $\text{MSD}(\tau) = \Gamma\tau^{\frac{1}{2}}$. For each time lag the mean was taken over 234 to 936 values.

Strain	$\sigma_\alpha^2\,[\mathrm{min}^2]$	$\sigma_\beta^2\,[\mathrm{min}^2]$	$\sigma_c^2\,[\mathrm{min}^2]$	$\sigma_{\alpha\beta}\,[\mathrm{min}^2]$	$\sigma_{\alpha c}\,[\mathrm{min}^2]$	$\sigma_{\beta c}\,[\mathrm{min}^2]$
N-P$_{cadBA}$	0.3600	0.0160	0.0230	0.0500	−0.0190	−0.0500
T-P$_{cadBA}$	0.0590	0.0320		−0.0330		
N+T-P$_{cadBA}$	0.0260	0.0500		−0.0300		

Table E.3: Covariance matrix. Covariance matrix of the parameters α, β and c from fitting the experimentally computed CDF to the sequential reversible model with mixed initial condition (N-P$_{cadBA}$) and fixed initial condition (T-P$_{cadBA}$ and N+T-P$_{cadBA}$).

E4 Spot Detection

After both the greyscale images and the fluorescence images were saved in tag image file format (tiff), the greyscale images were used to segment the cells with Oufti. The resulting Matlab files can be loaded together with the fluorescence microscopy images into our custom Matlab software *SpotDetection.m*. After testing the parameters and optionally adjusting the x and y values of the shift if the fluorescence images are shifted relative to the greyscale images, running the analysis saves the data of all detected fluorescence spots in all frames. An example image of the GUI of *SpotDetection.m* is shown in fig. E.3.

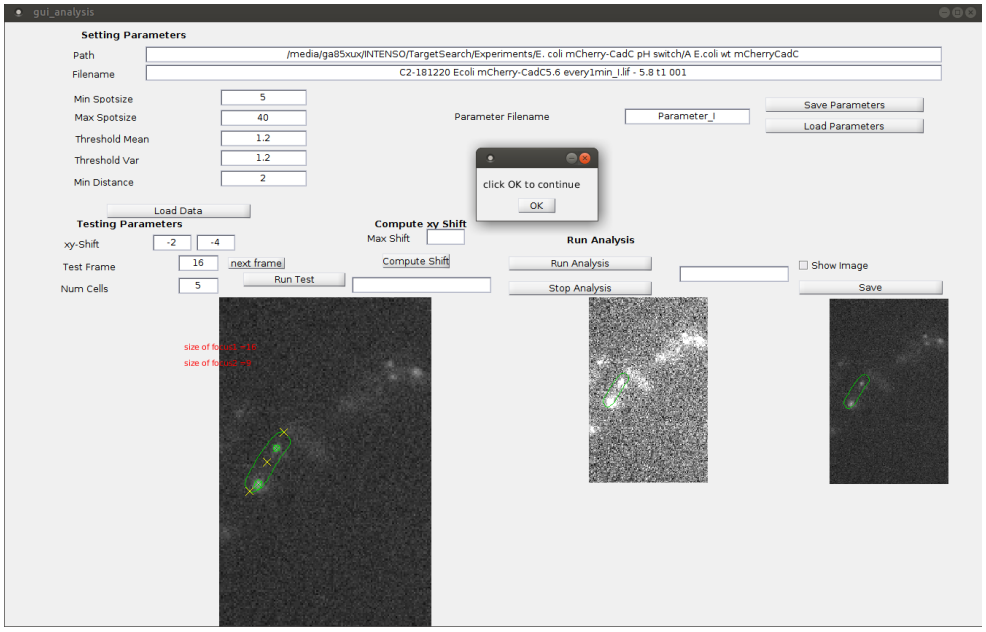


Figure E.3: *SpotDetection.m* software. Example image of the GUI of our custom Matlab software *SpotDetection.m* showing the test mode. To adjust the parameter settings the spot detection algorithm can be tested on a certain frame for a predefined number of cells. The graphical output shows the fluorescence image with detected spots of a certain size given in pixels, the cell outline, the position of the cell poles and the position of the cell centre. The values for the xy-shift can be adjusted if the fluorescence images are shifted relative to the greyscale images, which would be visible as a shift of the drawn cell outline relative to the cell in the fluorescence image.

Acknowledgements

The past years of pursuing a PhD have not always been easy. Sometimes I almost got lost in some detail, sometimes I started questioning the whole approach I was following. All the more I am deeply grateful to all people who have helped me getting back on track, both scientifically and mentally, such that I can finally write these lines that complete my dissertation.

First and foremost I would like to thank my supervisor Prof. Dr. Ulrich Gerland for giving me the chance to do my PhD with him. Thank you for giving me the freedom to find my own way but also for being there to discuss and give advice whenever I needed it. The friendly atmosphere you created helped me to become more confident in my decisions and grow both scientifically and as a person.

Moreover I would like to express my gratitude to Prof. Dr. Kirsten Jung, Dr. Sophie Brameyer and Elisabeth Hoyer for a very fruitful and interesting collaboration. Thank you Sophie and Elli for the great experimental work and for your patience with my ongoing questions about the biological details.

I would also like to thank my mentor Prof. Dr. Chase Broedersz for very useful discussions especially at the beginning of my PhD, your critical questions have helped a lot in clarifying my research question.

Furthermore I would like to thank all current and former members of the T37 group. From the very first day I have felt welcome in the group and the warm and friendly atmosphere helped a great deal to carry on with my PhD. Special thanks to Nanni for being a true friend, there is no problem an extended coffee break and a piece of Sicilian chocolate cannot fix. Thank you Elena for many fun memories and an awesome trip to Rome. Thank you Maryam, your cheerfulness has always brightened the sometimes grey office days. Thank you Vladi for many fun facts about Swabians combined with the matching bottle of wine. Thank you Hamid and Cesar for your help with my manuscript. Thank you Tobi and Joachim for always being there to discuss programming questions and problems with the cluster. Thank you Mareike for great shared work during the tutoring of “Stochastic Processes”. Thank you Flo and Jonas for being awesome office mates and providing a quiet working atmosphere. Thank you Manon for your help with the spot detection. And to all other group members not mentioned so far, thank you for making the group the way it is and for many fun memories of group retreats, conferences and Stammtisch nights.

I am very grateful for having been part of the TRR 174. Besides scientifically very interesting conferences and retreats it was a pleasure to meet many talented fellow PhD students and senior scientists and having the chance to discuss in a relaxed and inspiring environment.

I would like to especially acknowledge those who helped me proofreading my dissertation. Basti, Louis and Nanni you have helped me a lot in finalising my thesis.

Special thanks to the secretaries Laura, Claudia and Daniela for their helpfulness and continuous support. Thank you for keeping an overview over all formalities and for always being there for a chat in the kitchen.

Also thanks to the Campus Chor Garching for providing some change during my PhD. Thanks to all fellow Stammtisch mates, especially to Christiane also for the endless walks during Covid times.

Finally, I would like to thank all the important people who have led me to this point in life where I am now. First of all, thanks to my family for your unconditional love and support. Thank you Basti, Johann and David for shared struggles, happy moments and fun times during our physics studies and for being true friends for life. The last sentence I dedicate to David and Jonas, thank you for being part of my life, I love you dearly.