Max-Planck-Institut
für Kolloid- und Grenzflächenforschung

Mechanisms of biochemical reactions within crowded environments

Dissertation
von

David Gomez

Kumulative Dissertation
zur Erlangung der Würde des akademischen Grades
“doctor rerum naturalium”
(Dr. rer. nat)

in der Wissenschaftsdisziplin
Theoretische Biologische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

angefertigt am
Max-Planck-Institut für Kolloid- und Grenzflächenforschung

Potsdam, März 2016

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-94593
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-94593

En este mundo no necesitamos mas doctores, necesitamos mas personas...

To my Family

Declaration of Authorship

I, David Gomez, declare that this thesis titled: "Mechanisms of biochemical reactions within crowded environments" and the work presented in it is my own. I confirm that:

– This work was done wholly while in candidature for a research degree at Potsdam University.

– I have quoted the work of others.

– I have acknowledged all main sources of help.

Signed: _____

Date: _____

Zusammenfassung

Innerhalb einer Zelle, im Zytosol, entstehen und arbeiten sehr viele biologische Makromoleküle. Die Dichte dieser Moleküle ist sehr hoch und dieses ‘vollgestopfte’ Zytosol hat vielfältige Auswirkungen auf viele biologische Prozessen wie zum Beispiel Protein-Protein Interaktionen, Genregulation oder die Faltung von Proteinen. Der Ablauf von vielen biochemische Reaktionen in dieser Umgebung weicht von denen unter verdünnte Laborbedingungen ab. Um die Effekte dieses ‘makromolekularen Crowdings’ zu verstehen, wurde in den letzten Jahren bereits viel Mühe investiert.

In dieser Arbeit kombinieren wir verschiede Computermethoden, um die Wirkungen des ‘makromolekularen Crowdings’ auf biologische Prozesse besser zu verstehen. Zuerst schlagen wir ein Gittermodell vor, um damit die Effekte des ‘makromolekularen Crowdings’ auf enzymatische Reaktionen zu studieren. Damit stellen wir ein detailliertes Bild zusammen, wie Crowding die Assoziations- und Dissozotationsraten beeinflusst und wie verschiedene crowding-Effekte zusammen auf die Gleichgewichtskonstante wirken.

Weiterhin implementieren wir ein Gittermodell der ‘erleichterte Diffusion’. Unsere Ergebnisse zeigen, dass Hindernisse an der DNA die vereinfachte Diffusion beeinträchtigen. Das Ausmass dieser Wirkung hängt dabei von der Dynamik der Hindernisse an der DNA ab. Im dem Fall dass Crowder ausschließlich in der Lösung vorhanden sind, erhöhen sich unter bestimmten Bedingungen DNA-spezifische Bindungen.

Schließlich nutzten wir strukturbasierte Techniken um damit die Auswirkungen von Crowding auf die Faltung von Proteinen zu untersuchen. Wir fanden dabei, dass Polymer Crowder stärkere Wirkungen auf die Proteinstabilität haben als kugelförmige Crowder. Dieser Effekt verstärkte sich mit der Länge der untersuchten Polymere. Die Methoden die hier vorgeschlagen werden, sind generell anwendbar und können auch an deutlich komplexeren Systemen angewandt werden.

Abstract

The cell interior is a highly packed environment in which biological macromolecules evolve and function. This crowded media has effects in many biological processes such as protein-protein binding, gene regulation, and protein folding. Thus, biochemical reactions that take place in such crowded conditions differ from diluted test tube conditions, and a considerable effort has been invested in order to understand such differences. In this work, we combine different computationally tools to disentangle the effects of molecular crowding on biochemical processes. First, we propose a lattice model to study the implications of molecular crowding on enzymatic reactions. We provide a detailed picture of how crowding affects binding and unbinding events and how the separate effects of crowding on binding equilibrium act together. Then, we implement a lattice model to study the effects of molecular crowding on facilitated diffusion. We find that obstacles on the DNA impair facilitated diffusion. However, the extent of this effect depends on how dynamic obstacles are on the DNA. For the scenario in which crowders are only present in the bulk solution, we find that at some conditions presence of crowding agents can enhance specific-DNA binding. Finally, we make use of structure-based techniques to look at the impact of the presence of crowders on the folding a protein. We find that polymeric crowders have stronger effects on protein stability than spherical crowders. The strength of this effect increases as the polymeric crowders become longer. The methods we propose here are general and can also be applied to more complicated systems.

Contents

Chapter 1

Introduction

1.1 Introduction

The interior of a cell is a crowded environment in which different macromolecules, such as DNA, proteins, lipids and sugars, coexist together with highly organized macromolecular structures like the cytoskeleton. One example is the cytoplasm of the *Escherichia coli* bacteria, that can be occupied up to 40% by such macromolecules (1; 2). Moreover, *in vivo*, almost all proteins and other macromolecules belong to specific biochemical reaction networks, which are necessary for them to be active and perform a task in the cell. A great effort has been made to characterize the attractive interactions between the components of biochemical reaction networks (3; 4; 5). However, unavoidable repulsive interactions between macromolecules are typically not considered *in vitro* studies. In such studies, reactions take place under dilute solutions that drastically differ from *in vivo* systems (6). Nonetheless, an increased effort has been made to understand the effects of the presence of these macromolecules on biochemical processes, including protein-protein binding (7; 8), protein folding (9; 10; 11; 12), gene expression (13; 14; 15; 16; 17; 18) and enzymatic activity (4; 19; 20; 21).

Molecular crowding is ubiquitous within biological systems, and its effects, although complex, can be generic and described by two principal contributions: On the one hand, binding equilibria is typically shifted towards the bound state (6; 8). On the other, the presence of crowders hinders diffusion in the solution (22; 23; 24). Thus, diffusion-limited biochemical reactions that require an encounter between different reactants cannot take place in the limit of very high volume occupation.

In the following chapters, we will focus on the effects that molecular crowding has on three rather general biochemical processes: We start by considering the effects of excluded volume on enzymatic reactions. Here, we present different reactions that uncouple the effects on binding equilibria from the effects on diffusion. In addition, we consider a Michaelis-Menten-like reaction that combines both contributions (25). As an outcome, we see that different enzymes can exhibit different behaviors under different crowding conditions, as previously shown by Zhou *et. al.* (4).

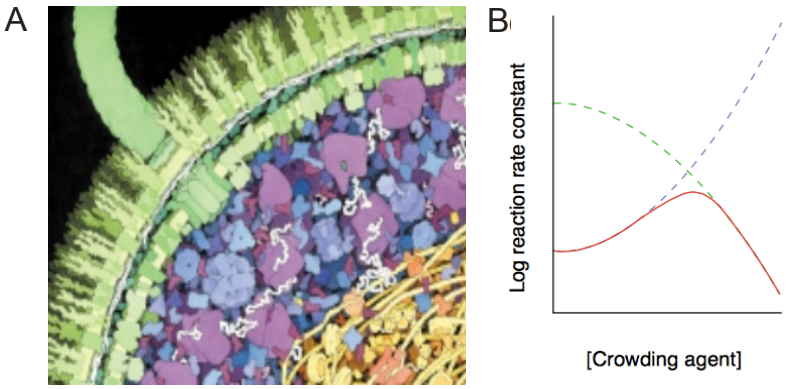

Figure 1.1 – A) Representation of the cell interior; from (6). B) overall reaction rate as a function of volume occupation, taken from (6). A non-monotonic behavior is obtained due to competition between the two opposite principal contributions of molecular crowding. The effects on the binding equilibria are in blue, and the effects on diffusion are depicted in green. The overall effect is shown as the red line with a maximum.

Next, we consider the effects of crowding on gene expression. The expression of a gene is a complex process that involves a series of processes and the orchestration of many molecular machines. More specifically, gene expression takes place when a specific DNA sequence of amino-acids is recognized by a DNA binding protein (DBP) (26; 27). Under crowding conditions, the specific recognition of a sequence can be bewildering. For example, if a transcription factor is non-specifically bound to the DNA, the presence of crowders can enhance non-specific binding, reducing, then, the average specific sequence finding time (18; 28). Moreover, many macromolecules bind to the DNA to perform functions such as DNA repair and initiation of transcription (29; 30; 31). Thus, the DNA can be highly occupied ($\sim 30\%$) by macromolecules. Under these conditions, the DBP's specific-sequence finding strategy has to be modified in order to be as efficient as possible (32; 33).

We finally discuss the effects that molecular crowding has on protein folding. Although a number of experimental and computational studies have been dedicated to understand the implications of excluded volume on protein folding (10), a systematic approach that takes different levels of volume occupation with different crowder geometries into consideration is lacking.

1.1.1 Molecular Crowding and Enzymatic Reactions

The cellular interior is a crowded environment that contains different kinds of macromolecules (34). There, the distance between macromolecules is comparable to the size of the macromolecule itself, see Fig. 1.1A. Thus, complex biochemical reaction networks, such as transcription, translation, protein-protein binding and protein folding, are expected to have different behaviors *in vivo* than in typically dilute test tube conditions (6; 7; 8).

In particular, the effects of molecular crowding on biochemical reactions can be divided into two principal components: On the one hand, molecular crowding shifts binding equilibria towards the bound state. This can be understood by the fact that once the associated state has been formed, more volume is available for the crowders. Thus, the total entropy of the associated system increases and the bound state becomes more stable (14; 35).

To provide a simple way of depicting the effect of molecular crowding on binding equilibria, let's consider a reaction of the type $A + B \rightleftharpoons AB$, where A and B are reactants. The latter reaction is a fully reversible reaction, and can be characterized by a dissociation constant $K_{\mathrm{d},0} = k_{\mathrm{off}}/k_{\mathrm{on}}$, where k_{off} is the unbinding rate, k_{on} is the reaction rate and the zero indicates an absence of crowders in the solution.

More generally, the reaction dissociation constant can be defined as $K_{\mathrm{d},\phi}$, where ϕ is the volume occupation fraction of the crowding agents. Thus, the effect on the dissociation constant is purely thermodynamic, and can be characterized by the statistic of the states in which the receptor is either free or occupied (5). Consider a 3-dimensional lattice with Ω lattice sites and a total volume $V = \Omega\mathfrak{l}^3$, where $\mathfrak{l}$ is the lattice spacing. Set a static target or receptor R, and L ligands and C crowders diffusing into the lattice with a given diffusion constant. If one ligand finds the receptor, the complex RL is form. At equilibrium, the probability for the complex to be form can be calculated as:

$$P_b = \frac{S_{\mathrm{b}} e^{-E_b/k_{\mathrm{B}}T}}{S_{\mathrm{b}} e^{-E_b/k_{\mathrm{B}}T} + S_{\mathrm{ub}}}, \tag{1.1}$$

where k_{B} is the Boltzmann constant, T is the temperature, S_{b} is the number of states in which the complex is formed, S_{ub} is the number of states in which the receptor is free, and E_b the binding energy. After obtaining S_{b} and S_{ub} (for more details please refer to Chapter 3, where the kinetics of this process will also be discussed in detail), P_b can be expressed as:

$$P_b = \frac{1}{1 + \frac{K_d(\phi)^r}{L/\Omega}}. \tag{1.2}$$

One can then show that the change in the dissociation constant as a function of ϕ is given by: $K_{\mathrm{d},\phi} = K_{\mathrm{d},0}(1-\phi)^r$, where r is a factor that depends on the crowder size (5; 25).

The second principal effect that molecular crowding exerts on biochemical reaction dynamics is the slowing down of diffusion. Different experiments at different volume occupation fractions ϕ were found to be described by the phenomenological expression:

$$D(\phi) = D(\phi = 0)(1-\phi)^{\kappa}, \tag{1.3}$$

where κ is a free fitting parameter (14; 22). Diffusion experiments in the presence of crowders for the protein carbon monoxide hemoglobin have shown that $\kappa = 0.36$ (14; 22).

Experimentally, the quantity typically measured is the association rate. It can be expressed as: $k_{\mathrm{on}}^{-1} = k_{\mathrm{r}}^{-1} + k_{\mathrm{diff}}^{-1}$, where k_{r} is the intrinsic association rate once the reactants are in contact, and k_{diff} is the rate at which the reactants get into close contact. Experimental results have shown that *in vivo* diffusion constants can be slow ($\approx 1\ \mu^2 s^{-1}$) (14; 32). Hence, the reaction is diffusion limited and the association rate is given by the classical Smoluchowski result $k_{\mathrm{on}}(\phi = 0) = 4\pi Da$, where a is the size of the target (36). Under crowding conditions, diffusion is affected, leading to an association rate given by: $k_{\mathrm{on}}(\phi) = k_{\mathrm{on}}(\phi = 0)(1 - \phi)^{\kappa}$.

The two principal effects of molecular crowding on enzymatic reactions are thus opposite and compete with each other: On the one hand, the shift of the binding equilibria towards the bound state enhances the overall reaction rate. On the other hand, as the system's excluded volume increases, the association rate decreases. Hence, the time the reactants need to first encounter becomes large and the overall reaction rate decreases. The latter effects lead to a non-monotonous behavior in the overall reaction rate as a function of the volume occupation fraction, and a maximum in the reaction rate is obtained, see Fig. 1.1B.

When performing experiments to study the effects of molecular crowding on enzymatic reactions, the use of 'good' crowders that do not interact with the reaction of interest is essential. Dextran and PEG have been shown to work as good crowding agents and as a matter of completeness, we refer to some experimental results documented in (4):

- Addition of PEG 6K increases the enzymatic activity of the protein AspP of the *E. coli* bacteria (37).
- Addition of dextran 10K enhances the formation of a defamer of bovine pancreatic trypsin inhibitor (38).
- Addition of Ficoll 70K and PEG reduces the second-order rate constant for diffusion-limited bimolecular association of beta-lactamase (TEM) and BLIP (39; 40; 41).
- Addition of BSA or ovalbumin increases the binding affinity of replication protein RepA for specific DNA sequences (42).
- Addition of dextran 70K, Ficoll 70K, or PEG 6K increases enzyme activity of isochorismate synthase (43).

1.1.2 Gene Expression

A key role in biology is the study of precesses associated with the so-called central dogma of molecular biology (5; 44; 45), one of the most important classes of processes in cellular life. These processes are responsible for the synthesis of polymer chains, such as nucleic acids and proteins that are at the heart of cellular life.

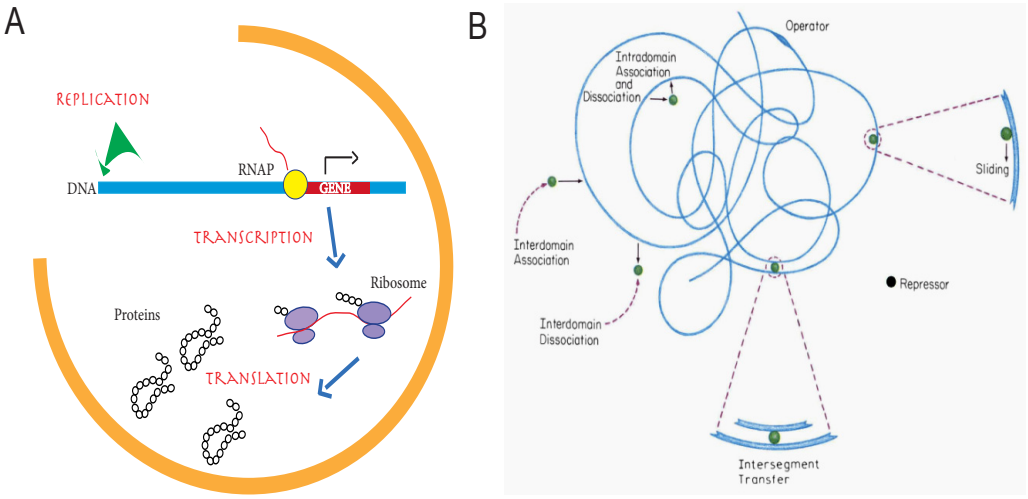

Figure 1.2 – A) Diagram of the central dogma of molecular biology. An RNA copy is generated from a DNA sequence after the RNA polymerase has translated a gene. The ribosome translates the RNA molecule and synthesizes proteins. B) Diagram of the facilitated diffusion process. A repressor follows a series of non-specific binding events, in which it slides along the DNA, until it finds its promoter. Figure taken from (70).

At the core of the central dogma is the DNA. This molecule encodes in nucleotide sequences, known as genes, all the RNA and protein products necessary for the cells. In its simple form, the central dogma of molecular biology can be presented as follows (5; 44): After the DNA has been replicated, a complex molecular machine known as the RNA polymerase (RNAP), transcribes the DNA genes and synthesizes an RNA molecule from it. More specifically, the RNAP synthesizes copies of messenger RNA (mRNA). This process is called transcription.

After an mRNA molecule has been synthesized, a second sophisticated molecular machine known as the ribosome, translates the mRNA content into amino acid chains that result in the central dogma's final product, proteins, see Fig. 1.2A.

The expression of a gene is a significantly more complicated process than the simplified version described above. Each gene is usually preceded by a regulatory sequence of nucleotides called the promoter. Different molecules can bind to this promoter. On the one hand, the RNAP can bind and start the transcription of the gene. On the other, regulatory proteins known as transcription factors, TFs, can also bind to the promoter. Depending on the nature of the TF, the outcome can differ drastically. If the transcription factor enhances the binding of the RNAP to the promoter, this TF is know as an activator, and the synthesis of an RNA molecule is enhanced. If, on the contrary, the TF hinders the rate at which the RNAP binds to the promoter, this TF is known as a repressor, and the expression of the gene is negatively affected (46). The interplay of different regulatory networks has been largely studied, and the results can be surprising (46; 47; 48; 49; 50; 51; 52; 53).

Interestingly, TFs exist in low numbers inside the cell (54). Different theoretical and experimental studies have shown that this issue has strong consequences on gene expression and how genes are regulated (55; 56; 57; 58). For example, it has been shown that a gene negatively regulated by a repressor reduces the variance of protein synthesis, even though protein synthesis is hindered (55; 59). Another key feature of gene expression is the fact that cells increase their volume during the cell cycle (60; 61; 62). Thus, TF concentration is not constant, changing as the cell volume increases. Computational and theoretical studies have shown that gene regulation can strongly differ from the classically considered constant volume scenario. For instance, the protein synthesis phase space of a positively regulated gene that exhibits bistability between a high and low protein synthesis rate, is affected depending on how large variations in volume increase are (51; 63; 64).

In addition to their low numbers, TFs' dynamics towards their promoters is a rather complicated process that can be sensitive to different conditions. In the following section, we present the facilitated diffusion model that TFs follow in order to find their specific base pair sequence on the DNA.

DBP-DNA binding and facilitated diffusion

Gene expression and regulation are mediated by transcription factors, or more generally, DNA-binding proteins (DBPs), that specifically bind to DNA sequences (65). This specific binding event can be a complex process, since the DBP needs to recognize a sequence that is only $\sim 10-30$ bp long (26; 66), from a genome that is, considering the bacteria *Escherichia Coli*, $\sim 10^6$ bp long (67). It is thus remarkable how quickly and precisely DBPs find their sequence target. In a seminal work, Riggs *et al.* measured the *in vitro* rate at which the *lac* repressor finds its promoter. They found that under certain conditions, the *lac* repressor binds to its target at faster rates than the simple three dimensional diffusion limit (68). As an explanation of this issue, Berg, von Hippel and Winter proposed the so-called facilitated diffusion model (69; 70). In this model, the DBP performs a series of three dimensional (3D) excursions in the bulk solution, together with one dimensional random walk events along the DNA, see Fig. 1.2B. This model has been supported by experimental results and is considered to be general to many DBPs (32; 71; 72; 73; 74).

In general, facilitated diffusion can be understood in the following way: Let D_1 and D_3 be the one-dimensional and three-dimensional diffusion constants, respectively. Experiments have shown that $D_3 > D_1$ (32; 71). Let $\epsilon_{\rm ns}$ be the non-specific DBP binding energy that corresponds to the free-energy difference between the protein being in the free solution and the protein being non-specifically bound to the DNA. Moreover, the DBP-DNA unbinding rate can be defined as $k_{\rm off} = A \exp\left(-\epsilon_{\rm ns}/k_{\rm B}T\right)$, where A is a constant, $k_{\rm B}$ the Boltzmann constant and T the temperature (75). At every non-specific binding event, the DBP scans on average $\lambda = (2D_1/k_{\rm off})^{1/2}$ base pairs. Thus, the effective size of the target is on average

$\lambda + a$, where a is the size of the target. If the DNA molecule has a length L, the DBP need on average $N = L/\lambda$ rounds of 1D sliding excursions and 3D free diffusion events in order to find the target. Following scaling arguments, the time the DBP spends on the DNA can be obtained as $t_{1D} \sim \lambda^2/D_1$, whereas the time the DBP spends in the bulk solution is given by $t_{3D} \sim V/LD_3$ (28; 76). Thus, the average time the DBP needs to find its target can be expressed as $t_{Tot} = N(t_{1D} + t_{3D})$:

$$t_{Tot} = B\frac{L\lambda}{D_1} + C\frac{V}{D_3}\lambda, \tag{1.4}$$

where B and C are dimensionless quantities that depend on the geometry of the system. A different argument will be presented in Chapter 4, where we study facilitated diffusion using on-lattice simulations. Depending on how large ϵ_{ns} is, the DBP experiences different dynamics: For very large values of ϵ_{ns}, k_{off} is very low and the DBP spends a long time on the DNA. Here, the first term of Eq. 1.4 dominates, and the DBP scans large sections of DNA. However, if the target is far from the DBP, it can spend too much time scanning regions where the target is not, so the process is inefficient. As ϵ_{ns} decreases, the unbinding rate k_{off} increases, and the DBP performs many series of 3D excursions and 1D sliding events (larger N), until the target is found. The latter makes the process more efficient and reduces the finding time. As ϵ_{ns} keeps increasing, the unbinding rate becomes so large that the DBP barely binds to the DNA. In this scenario, the DBP effectively looks for its target as in the free 3D diffusion case, and the finding time becomes large. Thus, the target finding time follows a non-monotonic behavior, in which there is an optimal ϵ_{ns} or k_{off} at which the dynamics is optimal. The latter is one of the hallmarks of facilitated diffusion, and has been experimentally and computationally confirmed (28; 32; 75; 77). In these approaches, the typical way to change non-specific binding energy is to modify the salt concentration. In this manner, the electrostatic interactions between the negatively charged DNA backbone and the positively charged receptor region of the DBP is screened (33; 78). Hence, very high salt concentrations imply a large k_{off} and vice versa.

Effects of molecular crowding on gene expression

Despite the fact that much theoretical (75; 76; 79; 80; 81; 82), experimental (66; 83; 84; 85; 86) and computational (28; 33; 75) effort has been spent in understanding the promoter finding process, it is still not fully understood. A particularly puzzling issue is the fact that cells are not empty. In fact, the cell interior is a crowded environment that can be occupied up to 40% by macromolecules (1; 87). In addition, many of these macromolecules bind to the DNA to perform functions such as transcription, DNA repair and gene regulation (26; 29; 30; 31). Consequently, the DNA is largely occupied $\sim 30\%$, by DBPs that affect the facilitated diffusion process.

In particular, Brackley *et al.* have shown that as the volume occupation ϕ increases in the bulk solution, DBP-DNA binding is enhanced and the protein remains bound for a longer time (28). This effect is complemented by a reduction in the time the DBP spends in the bulk solution. The latter effects modify the free facilitated diffusion case described by Marko *et al.*, in which for a given ϵ_{ns}, the optimal dynamics takes place when the time the DBP is bound to the DNA is equal to the time the DBP spends in the bulk solution (76). We note that the latter effect can depend on the DBP's initial conditions. Since diffusion is hindered by the presence of crowders, as ϕ increases, the DBP needs more and more time to its first encounter with the DNA. Thus, in the limit of very high excluded volume, the effect described by Brackley *et al.* may not be valid.

Additionally, different studies have shown the effects that obstacles on the DNA have on facilitated diffusion (28; 33; 88; 89; 90; 91). In general, they found obstacles on the DNA have a negative effect on facilitated diffusion, leading to larger specific-sequence finding times. As the DNA occupation fraction increases, the chances that the DBP will be enclosed between two obstacles increases (88). Thus, the sliding length of every 1D sliding event is determined by the average distance between the obstacles. The latter causes the DBP to spend more time finding its target, and to make more 3D excursions on average in order to overcome the presence of the obstacles (28; 88).

The DNA is a large sequence of base pairs that can have different configurations. It can be stretched and long or compact and coiled. For example, the *Escherichia coli* genome is $\sim 4.6^6$ base pairs long. With the help of nucleoid-associated proteins, this very long genome can be compacted into a small volume that occupies only about one-fourth of the intracellular volume (92). Experimental and theoretical studies have shown that for the protein EcoRV, a coiled DNA conformation enhances the specific sequence binding time (77; 93). Here, a coiled DNA conformation increases the intersegmental DBP binding events. That is, the DBP can jump from one region of the DNA to another, and keep scanning the DNA in 1D dimensional sliding events without completely unbinding from the DNA and going back to the bulk solution (77; 93).

1.1.3 Protein Folding

During translation, a polypeptide chain is synthesized as the ribosome decodes the information stored in the messenger RNA. If the polypeptide chain has a number n of residues, it can, in principle, fold into 8^n different conformations (45). The eight comes from the fact that only 8 bond angles are stereochemically allowed in the polypeptide backbone. Thus, taking the bacteria *Haemophilus influenzae*, which synthesizes proteins that are on average ~ 250 amino acids long (94), as an example, the average number of conformations will be on the order of $\sim 8^{250}$. This is a very large number. Nevertheless, in general, all proteins adopt a single 3-dimensional (3D) conformation, which is, under physiological conditions, the most

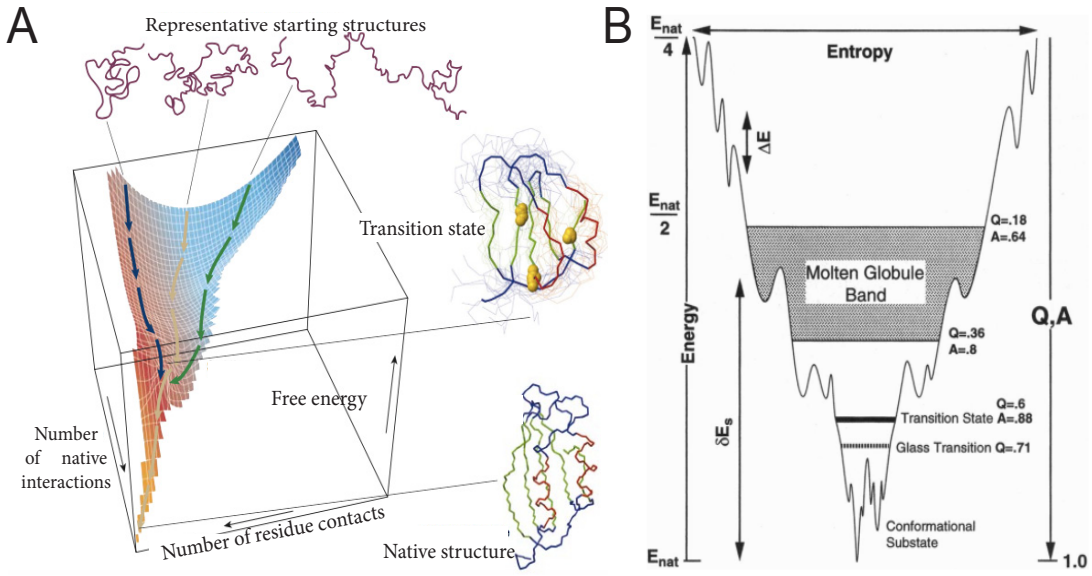

Figure 1.3 – Folding energy landscape. A) taken from (97) and B) taken from (101). The starting linear polypeptide chain follows a rugged funnel-shaped energy landscape. The native structure of the protein is located at the global minimum of the energy landscape.

stable of the possibilities. This folded structure is called the native state, and is required for protein activity (45; 95; 96; 97; 98).

This remarkable feature of proteins reveals that they are not simply a random sequence of polypeptides, but rather that they posses a defined compact active structure. But what guides the protein to the folded state? How is it possible that a one-dimensional chain leads to the folding of a particular protein? Much effort has been invested in answering these kinds of questions. For example, Kauzmann *et al.* quickly realized the great importance of hydrophobic forces on protein folding (99). These findings led to the analysis and development of random walks models, in which only two kinds of randomly distributed amino acids, hydrophobic and hydrophilic, are used to describe the formation of secondary structures in proteins (5; 95).

In the late 80s, a breakthrough in the understanding of protein folding started to be developed. It stated that in order to understand protein folding, what is needed is a global overview of the protein energy's landscape (95; 96; 100). The main idea is that the folding energy landscape resembles a rugged funnel, in which the protein can transiently stay at each local minimum, but eventually falls into the global minimum, reaching the protein native structure (95; 97), see Fig. 1.3.

The energy landscape model

A random heteropolymer chain folding in a lattice is one of the simplest models that can be understood by studying the energetic properties of the possible structures. The simplest case

of this model, is the one in which there are only two kinds of randomly distributed residues: hydrophobic and hydrophilic.

We start by defining the energy of a configuration α of the chain as: $E_\alpha = \sum_{\langle \mathrm{ab} \rangle} U_{\mathrm{ab}}$, where $\langle \mathrm{ab} \rangle$ means that the summation only goes on the non-polymeric nearest neighbor pairs, and U_{ab} is the interaction energy of the pair $\mathrm{a}-\mathrm{b}$. Now we define the partition function of the system as: $\sum e^{-U_\alpha / k_{\mathrm{B}} T}$, where k_{B} is the Boltzmann constant and T is the absolute temperature. The summation goes over all possible chain conformations of the polymer Ω_0.

If we take the limit of very large non-polymeric nearest neighbor pairs, in such a way that we can define N as the total number of interactions, and consider the interaction between pairs to be random variables, we see that the respective energies for polymer configurations are also random variables taken from a Gaussian distribution. Thus, we define the following relations:

$$\langle E_\alpha \rangle = N\epsilon_0, \quad \langle E_\alpha^2 \rangle = N\sigma^2 \quad \text{and} \quad p(E) = \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left(-\frac{(E - N\epsilon_0)^2}{2N\sigma^2} \right). \tag{1.5}$$

With the latter energy probability distribution, we can define the entropy $S(E)$ of the system: $S(E) = k_{\mathrm{B}} \ln(\Omega_0 p(E))$. As the system cools down, the entropy decays until the system runs out of entropy at a critical energy E_{C}. Thus, E_{C} is obtained as:

$$S(E_{\mathrm{C}}) = 0 \Rightarrow \ln(\Omega_0) - \frac{(E_{\mathrm{C}} - N\epsilon)^2}{2N\sigma^2} = 0. \tag{1.6}$$

It is also possible to obtain the corresponding critical temperature T_{F}, at which the system loses all its entropy. It is given by:

$$T_C^{-1} = \frac{dS(E)}{dE}\Big|_{E_{\mathrm{C}}} = -k_{\mathrm{B}} \frac{(E_{\mathrm{C}} - N\epsilon)}{N\sigma^2} = \sqrt{\frac{2k_{\mathrm{B}} S_0}{\langle E_\alpha^2 \rangle}}, \tag{1.7}$$

where we have used Eq. 1.6, and defined $S_0 = k_{\mathrm{B}} \ln(\Omega_0)$. For temperatures bellow T_C, the system freezes and behaves in a glassy-like manner (95). Although the state at the critical temperature can be thought of as the native state of the heteropolymer, it is not quite right, since at such T_C, the number of states are drastically decreased, and it is improbable that the native state of the polymer is the same state at such low temperatures. Moreover, at temperatures close to T_C, the polymer faces high energy barriers when trying to change its conformation. Thus, the polymer dynamics slows down, contradicting the idea of proteins folding quickly and easily (95; 101; 102; 103).

A more realistic approach is the one in which, with every native contact, local conformational energies stabilize the protein and the energy landscape follows a funnel-like shape with the native state at its global minimum (95; 101; 102). One useful parameter to describe the position of an ensemble of states in the funnel is the so-called Q value, which is the fraction of native-like contacts. At temperatures close to the folding temperature T_{F}, Q is close to 1,

and the protein can only take a few different conformational states, all of them similar to the native state. As the temperature increases, Q decreases and the ensemble of conformations increases.

To obtain a complete statistical mechanical description of the folding process, all thermodynamic variables that depend on the order parameter Q have to be determined. Using the approximation of no correlations among different energy states (95), we see that the energy of a random conformation is given by the contribution of many different random conformations, so the energy probability distribution is a Gaussian centered at the mean energy $\hat{E}(Q)$ and with variance $\sigma = \Delta E^2(Q)$:

$$p(E(Q)) = \frac{1}{\sqrt{2\pi\Delta E^2(Q)}} \exp\left(-\frac{(E-\hat{E}(Q))^2}{2\Delta E^2(Q)}\right). \tag{1.8}$$

With the number of possible configurations with a similar Q value as the native state $\Omega_0(Q)$, we define the entropy as:

$$S(E(Q)) = k_{\mathrm{B}} \ln(\Omega_0 p(E(Q)) = k_{\mathrm{B}} \ln(\Omega_0) - k_{\mathrm{B}} \frac{(E-\hat{E}(Q))^2}{2\Delta E^2(Q)}. \tag{1.9}$$

From the latter expression, the corresponding entropy for configurations similar to the native state can be defined as $S_0(Q) = k_{\mathrm{B}} \ln(\Omega_0)$.

At thermal equilibrium, only a small energy region is occupied by configurations similar to the native one. The most probable energy of this energy region can be found by maximizing the probability of the system to be in the native state. This is given by:

$$p(E(Q)) = \frac{\Omega(E) e^{-E_n/k_{\mathrm{B}}T}}{Z}, \tag{1.10}$$

where Z is the partition function that normalizes the probability function. After maximizing the latter equation, the most probable energy is given by:

$$E_{\mathrm{m.p}}(Q) = \hat{E}(Q) - \frac{\Delta E^2}{k_{\mathrm{B}}T}. \tag{1.11}$$

From Eq. 1.9, we obtain the number of thermally occupied states, and the entropy of the most probable energy:

$$\Omega(E_{\mathrm{m.p}}(Q)) = \exp\left(\frac{S_0(Q)}{k_{\mathrm{B}}} - \frac{\Delta E^2}{2(k_{\mathrm{B}}T)^2}\right), \tag{1.12}$$

$$S(E_{\mathrm{m.p}}(Q)) = k_{\mathrm{B}} \ln[\Omega(E_{\mathrm{m.p}}(Q))] = \frac{S_0(Q)}{k_{\mathrm{B}}} - \frac{\Delta E^2}{2k_{\mathrm{B}}T^2}. \tag{1.13}$$

With Eqs. 1.11 and 1.13, the free energy as a function of Q is defined at a given temperature T as: $F(Q,T) = E_{\mathrm{m.p}}(Q) - TS(E_{\mathrm{m.p}}(Q))$. If T is high, the free energy as a function of the

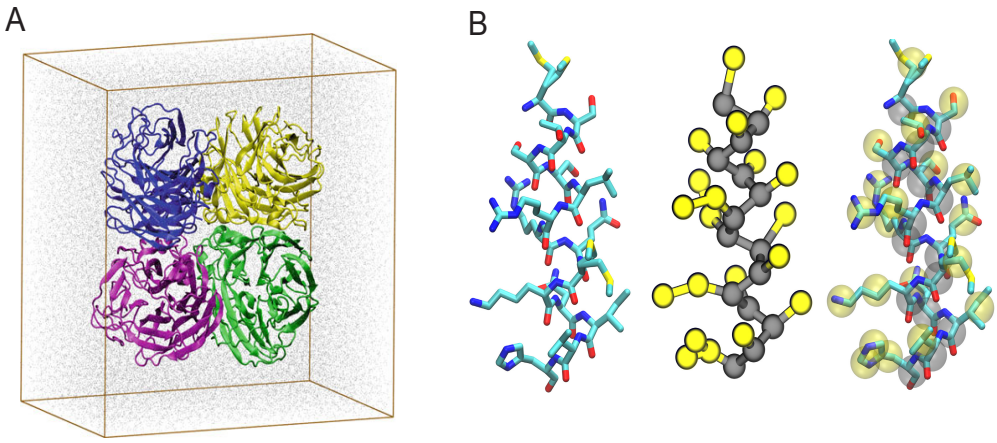

Figure 1.4 – A) MD simulation system for the avian influenza virus H5N1, taken from (107). B) Full-atom and coarse-grained representation of an example protein helix, taken from (109).

native-like contacts shows a minimum at low values of Q. This corresponds to an unfolded state. If T is low, the free energy exhibits a minimum at large values of Q. Thus representing a folded state. For intermediate temperatures, the free energy has two minima, corresponding to a folded and an unfolded state. At the folding temperature T_{F}, the probability of being in the folded state is the same as the probability of being in the unfolded state. Moreover, T_{F} defines a minimum Q value at which the two minima are equal. Thus, at T_{F} the following condition is fulfilled: $F_n = F(Q_{\min}, T_{\mathrm{F}})$. Taking as a first approximation that the entropy of the native state is close to zero, such that the free energy is equal to the internal energy E_{N}, the previous equality leads to an expression for the slope of the funnel:

$$\delta E_s / T_{\mathrm{F}} = S_0(Q) + \frac{\Delta E^2(Q_{\min})}{2k_{\mathrm{B}}T_{\mathrm{F}}^2}, \tag{1.14}$$

where $\delta E_s = \hat{E}(Q_{\min}) - E_{\mathrm{N}}$. Finally, we compare the previously obtained critical temperature T_C, in which the glass transition takes place, to the folding temperature F_{F}. Analytical and computational studies have shown that the ratio F_{F}/T_C can distinguish between fast and slow folding dynamics (95; 101). A schematic funnel for a realistic 60 amino acid long chain is depicted in Fig. 1.3B (101).

Molecular dynamics simulations for proteins

Experimental methods like X-ray crystallography provide a detailed picture of proteins in their native state. However, other experiments indicate that local motion inside the proteins is important for their functionality (104; 105; 106). Thus, in order to have a better understanding of proteins, it is essential to have detailed information about their dynamics, either during protein folding or protein activity. Molecular dynamics simulations provide the pos-

sibility to study relevant biochemical systems, that together with experimental results can give a multiscale description of the dynamics of biomolecules under study.

In a simple form, molecular dynamics (MD) is a computer simulation method that integrates the classical Newtonian equations of motion for a set of molecules, in order to obtain the trajectory of the system for a period of time (107). In Fig. 1.4A, we show a simulation system in which the enzyme neuraminidase from the avian influenza virus H5N1 is simulated with explicit water (107). For the integration of the Newtonian equations, the form of the potential energy U is of great importance. In MD, U is given as a force field, in which U has many components that characterize the different interactions in the biochemical system of interest. Moreover, in MD, different force fields with different sets of parameters are used depending on a particular application and on the set of parameters that fits the experimental data the best (108).

MD simulations in which all the atoms are explicitly simulated are known as all-atom simulations. If the simulated system is big, computational power becomes a constraint, and only small periods of time can be simulated. Thus, relevant biochemical reactions can no longer be observed, and the usefulness of MD becomes limited. To solve this issue, two different approaches can be used to reduce the computational cost: One is to reduce the number of simulated particles, and the other is to simplify the force field of the system. To reduce the number of simulated particles, simulations can be run with a group of atoms represented in a simplified model. This is the so-called coarse-grained implementation. Hence, a modified potential together with a smaller number of simulated particles decreases the computational cost, and longer simulation times can be obtained (109; 110). This leads to the acquisition of longer trajectories and the observation of slow biochemical reactions. Typically, proteins are simulated at the C_α level (110; 111; 112). That is, the amino acid residues are represented by only one particle centered at the C_α atom, which is the first carbon atom that attaches to a functional group, see Fig. 1.4B. Additionally, the complexity of the force field can be reduced under the framework of the funnel-like energy landscape presented above. Here, the models impose a native bias by explicitly including structural information in the force field. Thus, an initially unfolded sequence of amino acids will be quickly directed towards the native state of the system by attractive native interactions (95; 113). These kind of implementations are known as structure-based models (SBM) (114; 115). In Chapter 5, we use a coarse-grained SBM to study protein folding and protein stability within crowded environments.

Effects of molecular crowding on protein folding

Protein folding takes place inside cells, where many macromolecules occupy space. It is generally expected that interactions of excluded volume nature affect the folding and stability of proteins. Specifically, effects of molecular crowding on the binding equilibria can have profound effects on the protein folding energy landscape. Zhou showed that under moderate

crowding conditions, the native state is stabilized. Here, the energy barriers of the rugged funnel are reduced and the protein folding energy landscape is shifted towards the native state. Moreover, for high crowding conditions, the unfolded state is then stabilized (116).

In a computational study, Cheung *et al.* used a coarse-grained C_α model of the globular all-β-sheet WW domain, to show that as the levels of volume occupation increase, the folding temperature of the protein shifts towards higher temperatures (10). In other words, protein stability is enhanced when increasing levels of ϕ. They found that this effect can be explained by the compaction suffered by the protein when excluded volume increases. As ϕ increases, the protein is localized in smaller reaction volumes, thus, the protein compacts and its radius of gyration decreases (10). If the protein is more compact at any volume occupation $\phi \neq 0$, then the change in entropy follows the relation $\Delta S = S_\phi - S_0 < 0$, because some conformations are restricted.

Cheung *et al.* also found that the folding rate is enhanced at moderate levels of volume occupation. However, at high levels of volume occupation, the folding rate was shown to decrease. The latter effect is attributed to the restriction of conformational states necessary for protein folding caused by crowding (10). The effects of molecular crowding on protein folding can be hard to pin down, and a systematic study of the effect of different crowding agents on protein folding is still lacking.

As a matter of completeness, we refer to experimental studies on protein funding under crowding conditions reported in (4):

- Addition of dextran 30K stabilizes the molten globule conformation of apomyoglobin at pH 2 with respect to heat- and cold-induced unfolding (117).
- Refolding rate of Rd-apocyt b562 increases by 30% at 30°C and by 80% at 20°C in the presence of 85 g/L PEG 20K (118).
- Addition of PEG 4K increases the temperature for thermal denaturation of DNase I by more than 15°C (119).

1.2 Overview

The thesis is organized as a 'cumulative thesis' and consists of five chapters. Chapters 2-4 consist of three manuscripts, out of which two have been published (Chapters 2 and 3), and one is submitted (chapter 4). Chapter 5 contains additional unpublished results. A list of published and submitted manuscripts is provided below. Chapter 2 focuses on biochemical reactions under crowded conditions. Chapter 3 studies the facilitated diffusion dynamics with and without the presence of obstacles on the DNA. Chapter 4 presents the effects of the presence of molecular crowding in the bulk solution on the facilitated diffusion dynamics. In chapter 5, we explore the effects of molecular crowding on protein folding.

1.3 List of Publications and author contribution

The results of this study are presented in four different chapters of this thesis. The manuscripts that correspond to chapters 2, 3 and 4 of this thesis have been published or are currently under review in international peer-reviewed journals. Chapter 5 is a manuscript in preparation.

- D. Gomez and S. Klumpp, Biochemical reactions in crowded environments: Revisiting the effects of volume exclusion with simulations, Front. Phys. 3, 45 (2015).

  Author contributions. The project was planned by David Gomez and Stefan Klumpp. The simulations and data analysis were done by DG. The paper was written by DG and SK.

- D. Gomez and S. Klumpp, Facilitated diffusion in the presence of obstacles on the DNA. Phys. Chem. Chem. Phys. 18, 11184-11192 (2016).

  Author contributions. The project was planned by David Gomez and Stefan Klumpp. The simulations and data analysis were done by DG. The paper was written by DG and SK.

- D. Krepel, D. Gomez, S. Klumpp and Y. Levy. Mechanism of facilitated diffusion during DNA search in crowded environments. Submitted.

  Author contributions. Dana Krepel, David Gomez, Stefan Klumpp and Yaakov Levy planned research. DK performed coarse-grained MD simulations, DG performed lattice MC simulations. DK, DG and YL analyzed data and wrote the manuscript.

- D. Gomez and S. Klumpp. Effects of non-spherical crowders on protein folding. In preparation.

  Author contributions. The project was planned by David Gomez and Stefan Klumpp. The simulations and data analysis were done by DG. The manuscript was written by DG and SK.

Chapter 2

Biochemical reactions in crowded environments[1]

Abstract

Molecular crowding is ubiquitous within cells and affects many biological processes including protein-protein binding, enzyme activities and gene regulation. Here we revisit some generic effects of crowding using a combination of lattice simulations and reaction-diffusion simulations with the program ReaDDy. Specifically, we implement three reactions, simple binding, a diffusion-limited reaction and a reaction with Michaelis-Menten kinetics. Histograms of binding and unbinding times provide a detailed picture how crowding affects these reactions and how the separate effects of crowding on binding equilibrium and on diffusion act together. In addition, we discuss how crowding affects processes related to gene expression such as RNA polymerase-promoter binding and translation elongation.

2.1 Introduction

The interior of cells is a crowded environment, quite different from the dilute solutions usually studied *in vitro* (2; 6; 7). For example, in bacterial cells, macromolecules can occupy up to 40 % of the volume during phases of rapid growth (1), and the water content can drop far below this level upon exposure to increased osmotic pressure (87). The importance of molecular crowding for understanding processes in cells is increasingly appreciated. Its consequences have been studied extensively in the context of protein-protein binding (7; 8), protein folding (9; 10; 11; 19), enzyme activity (4; 20; 21), and gene regulation (13; 14; 15; 16; 17). Beyond these fundamental aspects, crowding has direct consequences to understand drought-

[1] Full, published title of the paper: Biochemical reactions in crowded environments: Revisiting the effects of volume exclusion with simulations

tolerance of plants (120; 121) and possibly neurodegenerative diseases in humans that are based on protein aggregation (122; 123).

Molecular crowding is ubiquitous, and the complexity of its direct and indirect effects can be bewildering. Some effects of crowding are generic (6; 7). For example, binding equilibria are typically shifted towards the bound state (7; 8; 124) and diffusion is slowed down (22; 23; 24). In the case of enzymes, both effects apply to the binding of the substrate to the enzyme, with opposite consequences. In addition, the reaction rate may be affected by specific changes in the molecular configuration of the enzyme that are induced by crowding or, in the case of multi-subunit enzymes, by increased binding between the subunits. As a result, different enzymes can exhibit rather different behavior upon increased crowding, as illustrated by the tabulated collection of experimental results in ref. (4).

In this article, we address the generic effects of crowding. We use a simple lattice model to disentangle the different contributions to the effects of crowding due to binding equilibria and diffusion as well as the dependence on the particle size. We specifically study molecular binding and enzymatic reactions. As a complement, we perform simulations with the reaction-diffusion simulation package ReaDDy (125). On the one hand, our approach is tutorial in nature and provides a simple rationalization and illustration of well-studied effects. On the other hand, it also provides some new insight. For example, in the case of a receptor-ligand pair, histograms of the binding times allow us to reconcile the reduced diffusion rate with the surprisingly crowding-independent binding rate.

Even though the effects we consider are generic, our main interest here is in gene expression, in particular reaction involving the macromolecular machines that process the genetic information, RNA polymerase and ribosomes. These machines are rather large molecular complexes, and for molecules of such size diffusion is strongly reduced (126). As a consequence, reactions are expected to be diffusion-limited or close to the diffusion limit. Specifically for ribosomes, it was recently proposed that the slow diffusion of ternary complexes (tRNAs charged with amino acids and GTP-activated elongation factor Tu) imposes a fundamental limitation on the speed of translation, which necessitates the large concentrations of elongation factors in rapidly growing bacteria (127) (elongation factor Tu is the most abundant protein in E. coli (128)). Such a limitation would be aggravated during growth under increased osmotic pressure. Likewise, RNA polymerase is a big molecular machine that diffuses slowly in the cell, such that binding to promoters, which usually determines the rate of transcription, might become diffusion limited.

The paper is organized as follows: We start by introducing the computation methods we use, the lattice model and ReaDDy. Then we study simple implementations of two elementary reactions, binding between two binding partners (which could represent two proteins or molecular complexes or a protein and its binding site on DNA) in section 3.3 and a diffusion-limited enzymatic reaction in section 3.4. Both cases are based on a ligand or substate that

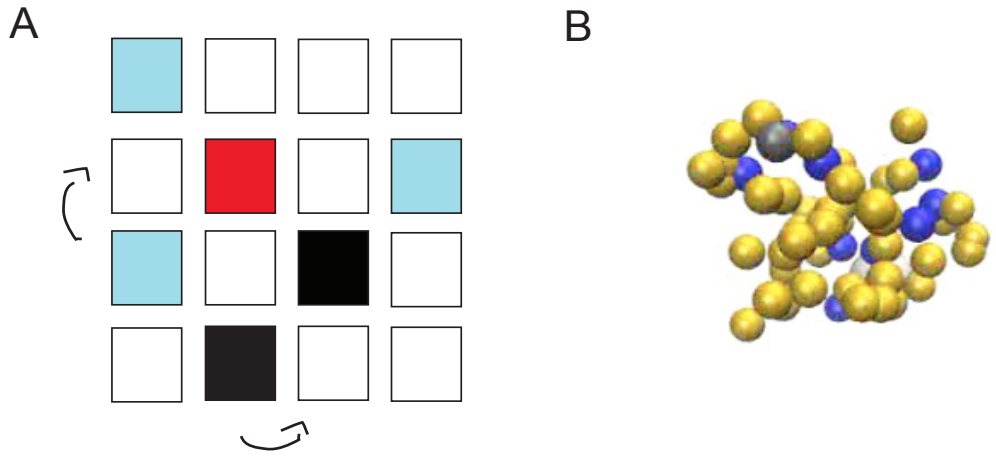

Figure 2.1 – Simulation methods used in this work: (A) Schematic view of the lattice implementation, in which L ligands (light blue) and C crowders (black) diffuse by hopping to one of the six nearest neighbor lattice sites until a ligand finds a target (red). Molecules have square geometry and diffusion is only allowed if the site of destination is unoccupied. Note that the simulations use a three-dimensional lattice rather than the two-dimensional one depicted schematically here. (B) Snapshot from the off-lattice simulation package ReaDDy. L ligands (dark blue), a single target (grey) and C crowders (yellow) diffuse and react in a simulation box with volume V.

needs to find a target site (the binding partner or enzyme) by diffusion. In section 2.5, we study the effects of the size of the crowders and in section 2.6 we combine everything to address an enzymatic reaction with Michaelis-Menten kinetics. We close with an extended discussion, where we apply the insight from these simulations to several processes relevant to gene expression, in particular to the speed of translation and to promoter finding by RNA polymerase.

2.2 Simulation methods

2.2.1 Lattice model

The effects of molecular crowding on biochemical reactions was studied using Monte Carlo simulations of particles on a three-dimensional lattice with periodic boundary conditions. The simulation box has the volume $V = \Omega l^3$, where l is the lattice spacing and Ω the total number of lattice sites, chosen as the linear extension of the smallest particle type. The system contains three types of particles: Target particles (receptors or enzymes), particles searching for the target (ligands or substrates, respectively) and crowders. Their numbers are denoted as R, L and C, respectively. For now, all particles are taken to occupy exactly one lattice site. We will consider particles of different sizes below, but throughout this study, all molecules on our lattice simulations are taken to have a square geometry. We will consider a single target particle (receptor or enzyme) and take this particle to be static in the center of

the simulation box. A ligand bound to the target particle is taken to occupy the same lattice site as the target particle. The other particles are initially placed randomly on the lattice, occupying the volume fraction $\phi \approx (R + L)/\Omega$, see Fig. 2.1A.

At each simulation time step (of duration τ), the crowders and ligands move to each neighbor site with probability $1/6$. This move is accepted if the chosen neighboring lattice site is free and rejected if it is occupied. Thus, steric effects or excluded volume are the only interactions considered, and on a free lattice, these particles diffuse with diffusion constant $D = l^2/(6\tau)$. These moves are performed in a random-sequential fashion: In every simulation step, we randomly choose L times a ligand and C times a crowder and update their position, thus that on average all particles are updated once per simulation step. If a ligand finds the target and the target is free, the complex ligand-target is always formed. Thus, this reaction is taken as diffusion-limited. Only when the target is already occupied by another ligand, the move to form the complex is rejected. Once the complex has been formed, the bound ligand can dissociate from the target with the unbinding rate k_{ub}. Below, we will consider different scenarios of complex dissociation to describe receptor-ligand complexes and enzyme-substrate reactions.

2.2.2 Off-lattice simulations (ReaDDy)

In addition to the lattice implementation, we run simulations with the off-lattice simulation software ReaDDy (125), see Fig. 2.1B. This simulation package describes biochemical reaction-diffusion processes via interacting spherical molecules. Their diffusion is described by a memoryless Langevin equation, which is solved numerically with an Euler discretization with a constant time step Δt,

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) - D\Delta t \frac{\nabla V(\mathbf{x}(t))}{k_B T} + \sqrt{2D\Delta t}\eta_t. \tag{2.1}$$

Here, $\mathbf{x}(t)$ is a three-dimensional vector indicating particle positions at time t, D is the reactant diffusion constant, $V(\mathbf{x}(t))$ is the particles' interaction potential, k_B is Boltzmann's constant, T is the temperature and η is a normally distributed random number with zero mean and variance one.

The steric interactions implemented in ReaDDy are given by a harmonic potential, in which the potential force constant k_{pot} is optimized by finding the largest simulation time step Δt for which there is no overlapping between particles (125). Throughout this study we use the recommended potential force constant $k_{pot} = 50$ kJmol^{-1} nm^{-1} (125). In ReaDDy, reactions are understood as uni- or bimolecular reaction events in which particles either are transformed into other particles, or events that lead to molecule synthesis or degradation. Thus, the concentration of molecules as a function of time can be described by the following set of ordinary differential equations (129):

$$\frac{dC_{P1}(t)}{dt} = ... = \frac{dC_{Pn}(t)}{dt} = k^1 C_R(t). \tag{2.2}$$

for the first order molecular reactions and

$$\frac{dC_{P1}(t)}{dt} = ... = \frac{dC_{Pn}(t)}{dt} = k^2 C_{R_1}(t) C_{R_2}(t), \tag{2.3}$$

for the second order molecular reactions. In the latter expressions, $C_{R_n}(t)$ and $C_{P_n}(t)$ are the concentration of the nth reactant and product species, and k^1 and k^2 are the first and second order reaction rates, respectively. We note that the macroscopic second order reaction rate k^2 is the product of the probability for the reactants to be in close contact with the rate at which the reactants in contact transform into the product, k^2_{micro}. Since diffusion is explicitly simulated by ReaDDy, only k^2_{micro} is set as an input in the simulations. For all simulations, we consider $k^2_{\text{micro}} = 10^8$ sec^{-1}. Computationally, a reaction takes place with a reaction probability p_{rea}, obtained from the Poisson probability of having at least one reaction event with rate $k^{1,2}_{micro}$ within time interval Δt, e.g. $p_{rea} = 1 - e^{k^{1,2}_{micro}\Delta t}$. We note that the molecular diffusion constant D and the reaction rates k^1 and k^2 are set in ReaDDy assuming dilute conditions, e.g. $\phi = 0$. By increasing levels of molecular crowding the kinetics and thermodynamics of the reactions are influenced.

To compare the results from both types of simulations, the length and time units of the lattice simulations must be converted to nanometers and seconds. In ReaDDy, we run simulations with spherical particles of radius $r = 3$ nm that diffuse within a solid square lattice of volume $V = 49 \times 49 \times 49$ nm^3 $= 1.13 \times 10^5$ nm^3. This volume is chosen in such a way that the total volume is 1000 times the molecules' volume, as in the lattice simulation. This choice defines our lattice in the $10 \times 10 \times 10$ lattice constant to correspond to $l = 4.9$ nm. The time scale τ of the lattice simulations then corresponds to $l^2/(6D)$, which can be used to convert the reaction rates. However, a small difference remains between the capture areas from which binding occurs, so that the binding probabilities P_b are not exactly the same in both methods. To correct for that difference, we adjust the unbinding rate, such that in the absence of crowding the binding probabilities agree between the two methods. This adjustment corresponds to an approximately two-fold increase of the unbinding rate, which thus corrects the corresponding two-fold increase of the binding rate due to different 'target volumes'. This adjustment procedure can be viewed as an instance of renormalizing the binding binding/unbinding rate by including unbinding events that are too short for the particle to diffuse away in the bound state (130).

2.3 Effects of crowding on molecular binding

We start by considering the simple case of binding between a receptor and a ligand in the presence of crowders. We emphasize that we consider a rather generic scenario here, where

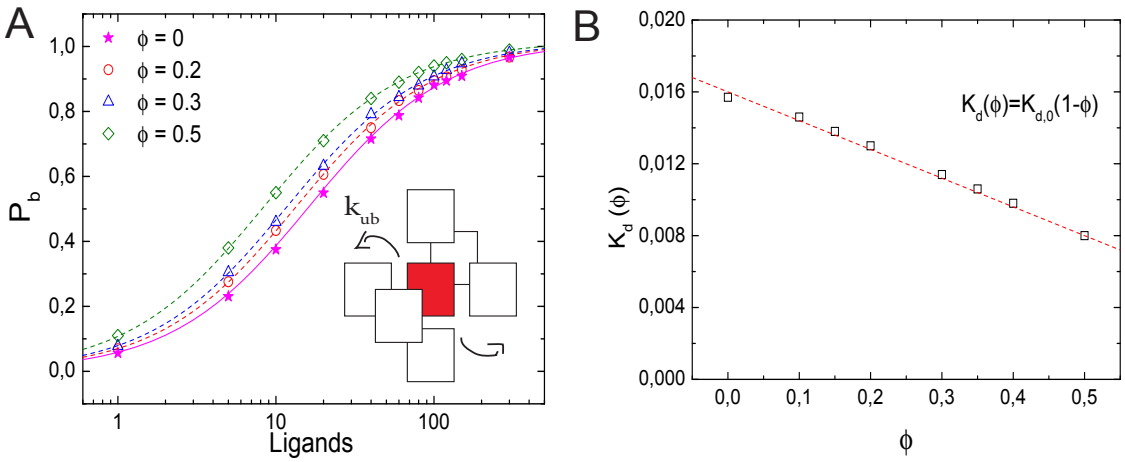

Figure 2.2 – Effects of molecular crowding on binding equilibria. (A) Probability for the receptor to be occupied, P_b, as a function of the number of ligands for four different values of volume occupation, $\phi = 0, 0.2, 0.3$ and 0.5. Crowding shifts the curves towards lower ligand concentrations. Simulation data are in agreement with Eq. 2.6 (lines). (Inset) Representation of the unbinding reaction. The ligand steps from the target to one of the six nearest neighbor sites with a rate k_{ub}. (B) Dissociation constant $K_d(\phi)$ as a function of ϕ. The data points are obtained from the simulations by interpolation (see text) and the line is $K_d(\phi) = K_{d,0}(1-\phi)$. Simulation parameters: $V = 10\text{x}10\text{x}10\, l^3$, $D = 1/6\, l^2\tau^{-1}$ and $K_{d,0} = 1/60$.

the receptor and ligand do not necessarily describe the typical case of a protein receptor and a small molecule ligand, but could, for example, also correspond to two proteins, to a regulatory binding site on DNA and a transcription factor, or even to an enzyme and its substrate (provided the actual reaction is very slow, as we will discuss below). Effects of excluded volume on such a reaction have been studied extensively in the past; in particular, it is well-known that crowding shifts the binding equilibrium towards the bound state (6; 7; 8). The lattice model allows us to provide a rather intuitive explanation of these effects. In addition, we use our simulations to investigate the effects of crowding on the kinetics, which are more subtle.

2.3.1 Binding equilibrium

In the simulations of receptor-ligand binding, a ligand reaching the target particle (receptor) forms the receptor-ligand complex unless the receptor is already occupied by another ligand. Thus, the binding reaction is diffusion-limited (a reaction-limitation could be introduced by accepting the binding move with a probability smaller than one but we do not consider this case here). The complex can dissociate again with a rate k_{ub}. Unbinding is implemented by randomly choosing a neighbor site of the complex and moving the ligand there, provided that site is free. At the same time, the receptor becomes free again and available for another binding event. Thus, the binding reaction, $L + R \rightleftharpoons LR$, is fully reversible and can be characterized by the dissociation constant $K_{d,0}$, where the index zero indicates the absence of crowder molecules.

Fig. 2.2A shows results from simulations of a lattice with volume $V = 10 \times 10 \times 10\ l^3$, i.e. with $\Omega = 1000$ lattice sites, where crowders, ligands and the receptor all occupy one lattice site each. In dilute conditions, the unbinding event is set to take place (unless otherwise stated) at a rate $k_{ub}(0) = 1/60\ \tau^{-1}$. To ensure equilibration, the simulations are run until 10000 binding events have taken place. Fig. 2.2A shows the probability P_b that the receptor is occupied, as a function of the number of ligands for different values of volume occupation ϕ. P_b is determined as the ratio of the time the receptor was occupied and the total simulation time. The symbols represent our simulation results for $\phi = 0, 0.2, 0.3$ and 0.5, the lines indicate the corresponding results from the binding equilibrium condition (no free parameters, discussed below), $P_b = C/(C + K_d(\phi))$, where the dissociation constant is expressed in units of numbers of molecules in the simulation box. Although qualitatively the behavior is similar for all values of ϕ, increasing volume occupation increases the receptor occupancy by shifting the curves to the left. Thus, as ϕ increases, less ligands are needed to saturate the receptor. For large numbers of ligands, the effect of the crowders is negligible, because the receptor is already saturated. In Fig. 2.2B we plot the dissociation constant $K_d(\phi)$, which is obtained by interpolation from the simulations as the ligand concentration for which $P_b = 0.5$, as a function of the volume occupation ϕ. $K_d(\phi)$ decreases linearly with the volume fraction, indicating the shift of the equilibrium towards the bound state.

Since the effect of crowders and excluded volume on the dissociation constant is purely thermodynamic rather than kinetic, it can be understood based on the statistics of states of the lattice with the receptor free and occupied, see ref. (5). We include this argument here for completeness. At equilibrium, the probability for the receptor to be occupied P_b, is obtained as the ratio of the number of possible states S_b, in which a ligand is bound to the receptor and the total number of states, i.e. S_b plus the number of states when the receptor is free, S_{ub}. We note that the states with an occupied receptor are weighted with the Boltzmann factor due to the binding energy E_b,

$$P_b = \frac{S_b e^{-E_b/k_{\mathrm{B}}T}}{S_b e^{-E_b/k_{\mathrm{B}}T} + S_{ub}} = \frac{1}{1 + \frac{S_{nb}}{S_b} e^{E_b/k_{\mathrm{B}}T}}. \tag{2.4}$$

The ratio S_{nb}/S_b is obtained by dividing all the possible ways in which C crowders and L ligands can be organized in a lattice with Ω lattice sites, over all the possible ways in which C crowders and $(L-1)$ ligands can be distributed in Ω lattice sites. This ratio is thus given by

$$\frac{S_{ub}}{S_b} = \frac{\binom{\Omega}{L+C}\binom{L+C}{L}}{\binom{\Omega}{L+C-1}\binom{L+C-1}{L}} \approx \frac{\Omega}{L}\left(1 - \frac{(L+C)}{\Omega}\right) = \frac{(1-\phi)}{L/\Omega}, \tag{2.5}$$

with L/Ω being the dimensionless ligand concentration. Introducing the latter expression into Eq. 2.4 leads to

$$P_b = \frac{1}{1 + \frac{K_d(\phi)}{L/\Omega}}, \tag{2.6}$$

with the crowding-dependent dissociation constant $K_d(\phi) = e^{E_b/k_\mathrm{B}T}(1-\phi)$. The limiting value for $\phi = 0$ is the microscopic dissociation constant for the binding reaction without interfering crowders, $K_{d,0} = e^{E_b/k_\mathrm{B}T}$. We note that this dissociation constant as well as the ligand concentration L/Ω are dimensionless, but can be converted to per-volume units as $[L] = L/V = l^{-3} \times L/\Omega$ and likewise for the dissociation constants. The microscopic dissociation constant $K_{d,0}$ can also be related to the kinetic parameters of the model via the detailed balance condition, $k_{ub}/\tau^{-1} = e^{E_b/k_\mathrm{B}T}$. The latter relation is used to determine the lines in Fig. 2.2A and B.

Eq. 2.6 thus shows that, as ϕ increases, the dissociation constant is diminished and a lower ligand concentration is needed to saturate the receptor as observed in the simulations. An alternative interpretation of the result is that the volume is reduced by the excluded volume, so that the available volume is $\Omega(1-\phi)l^3$, which effectively increases the ligand concentration. We note however that this interpretation is only valid in the simplest case that we consider here, as it depends on the assumption that all particles have the same size. In that case, the available volume is independent of the spatial arrangement of the particles, which is not the case for particles of different sizes, as discussed below.

2.3.2 Kinetics of binding and unbinding

Next we consider the impact of the crowders on the kinetics of binding and unbinding. To that end, we determine binding and unbinding rates from our simulations as the inverse of the average time the receptor is free before a binding event and occupied before an unbinding event. In Fig. 2.3A we plot the binding rate k_b for three different ligand concentrations as a function of the occupied volume fraction. The binding rate is given by $1\, l^3 L^{-1} \tau^{-1}$ and depends on the ligand concentration, but not the level of crowding, suggesting that diffusion of the ligand to the target is not affected by crowding. The observation of a constant binding rate indicates that the effect of crowding on the dissociation constant discussed above is entirely due to the decrease of the unbinding rate $k_{ub}(\phi)$. In Fig. 2.3B we plot $k_{ub}(\phi)$ as a function of the excluded volume fraction ϕ. As expected, $k_{ub}(\phi)$ decreases linearly as the level of ϕ increases, consistent with the expression $k_{ub} = K_{d,0}\tau^{-1}(1-\phi)$. Such a linear decrease can be understood as a simple exclusion rule for the complex dissociation step, where the ligand moves from the receptor site to one of its neighbors. The rate for that step is simply reduced by a factor corresponding to the probability that this site is free, $(1-\phi)$. We can thus conclude that for the simple case with crowders and ligands of the same size, the shift of the binding equilibrium towards the bound state is caused by crowders blocking unbinding.

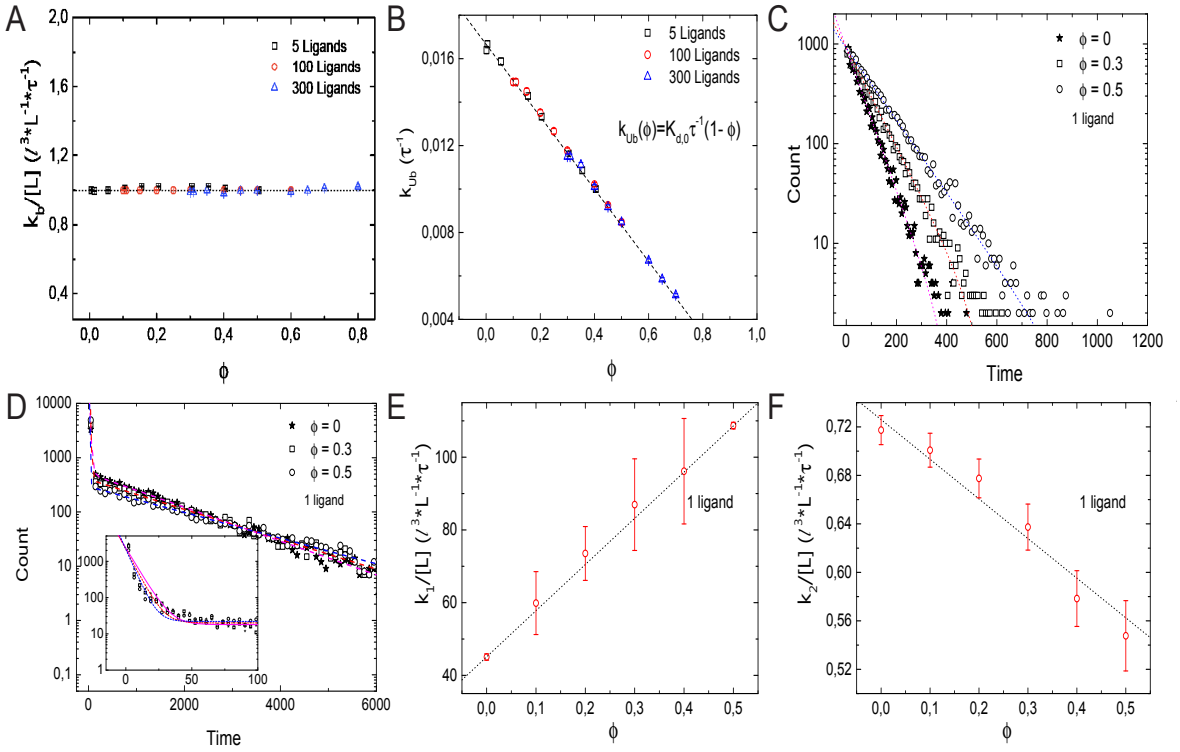

Figure 2.3 – Kinetics of the simple binding reaction model. (A) Binding rate k_b and (B) unbinding rate k_{ub}, as a function of the volume occupation fraction ϕ. k_b remains constant for all values of ϕ. Thus, diffusion of the ligand towards the target is not affected by crowding. k_{ub} decays as ϕ increases, in agreement with the expression $k_{ub} = K_{d,0}\tau^{-1}(1-\phi)$. (C) Histograms of the duration of 10000 unbinding events for the case of a single ligand for different values of ϕ. The data follow a single exponential. (D) Histograms of the duration of 10000 binding events for $\phi = 0, 0.3, 0.5$ and a single ligand. A double exponential decay is observed with a fast and slow component in the binding kinetics. Inset: closeup of the histogram for short times with smaller bin size. (E) Fast rebinding is enhanced when levels of molecular crowding increase. (F) Slow rebinding is negatively affected when ϕ increases, since the diffusion of the ligand towards the target is hindered. Simulation parameters as in Fig. 2.2.

The rates presented so far are based on the mean values of times between binding and unbinding events. To get a more detailed picture of the effects of crowders, we next consider histograms of these times. Fig 2.3C shows the histograms of the durations of the bound state before unbinding for the case of a single ligand and three different volume occupation values. The simulation data follow a single exponential (correlation coefficient of fit $R^2 = 0.984, 0.995$ and 0.998 for $\phi = 0, 0.2$ and 0.5, respectively), with a characteristic time scale that increases as the volume occupation increases, in agreement with an unbinding rate reduced by crowding. For the binding times, the picture is more complex: Fig 2.3D shows the histograms of the durations of all binding events for the same three cases. In all cases, the simulation data can be described by a double exponential (correlation coefficient of fit $R^2 = 0.999$ for the three values of $\phi = 0, 0.2$ and 0.5), $P(t) = N_1 \exp(-k_1 t) + N_2 \exp(-k_2 t)$, which indicates the existence of a fast and a slow component in the binding kinetics. The corresponding two rates k_1 and k_2 are plotted individually as functions of the volume fraction in Figs. 2.3E and 2.3F, respectively. These plots show that they exhibit opposite dependencies on crowding. The rate for rapid rebinding, k_1, increases with increasing volume fraction ϕ, while the rate of the slow component decreases. This result can be interpreted in the following way: crowding enhances rapid rebinding of a ligand still close to the receptor upon unbinding, as the crowders hinder the diffusive motion of the ligand away from the receptor and thus keep it close for a longer time. At the same time, the crowders hinder the diffusive motion to the receptor when the ligand needs to diffuse there from further away. Thus, the constant binding rate, independent of the presence and concentration of crowders conceals the more subtle balance between two opposing effects, namely an increased rate for rapid rebinding and a reduced rate for binding during longer periods of unoccupied receptor. Both effects are related to an effect of crowding on diffusion, but in one case away from the target and in the other towards the target. One could however, consider coarse-grained rates and consider the short-lived unbinding events as part of the bound state (130). In that case, the binding rate would be reduced by crowding and the unbinding rate would be reduced even more.

2.3.3 Binding in off-lattice simulations

In addition, we simulated the same receptor-ligand binding reaction using the off-lattice reaction-diffusion dynamics software ReaDDy (125). These simulations allow us to cross-check the results and to test for possible lattice artifacts. Fig. 2.4A shows the probability P_b that the target is occupied as a function of the number of ligands for different values of ϕ. Comparable quantitative results are found for both implementations of molecular crowding. As in the lattice simulations, increasing volume occupation levels ϕ leads to a decrease in the dissociation constant. In Fig. 2.4B and 2.4C we show the rates of binding and unbinding as functions of the volume occupation. The binding rate is independent of the volume fraction, while the unbinding rate decreases as the volume fraction of crowders increases. Just as in

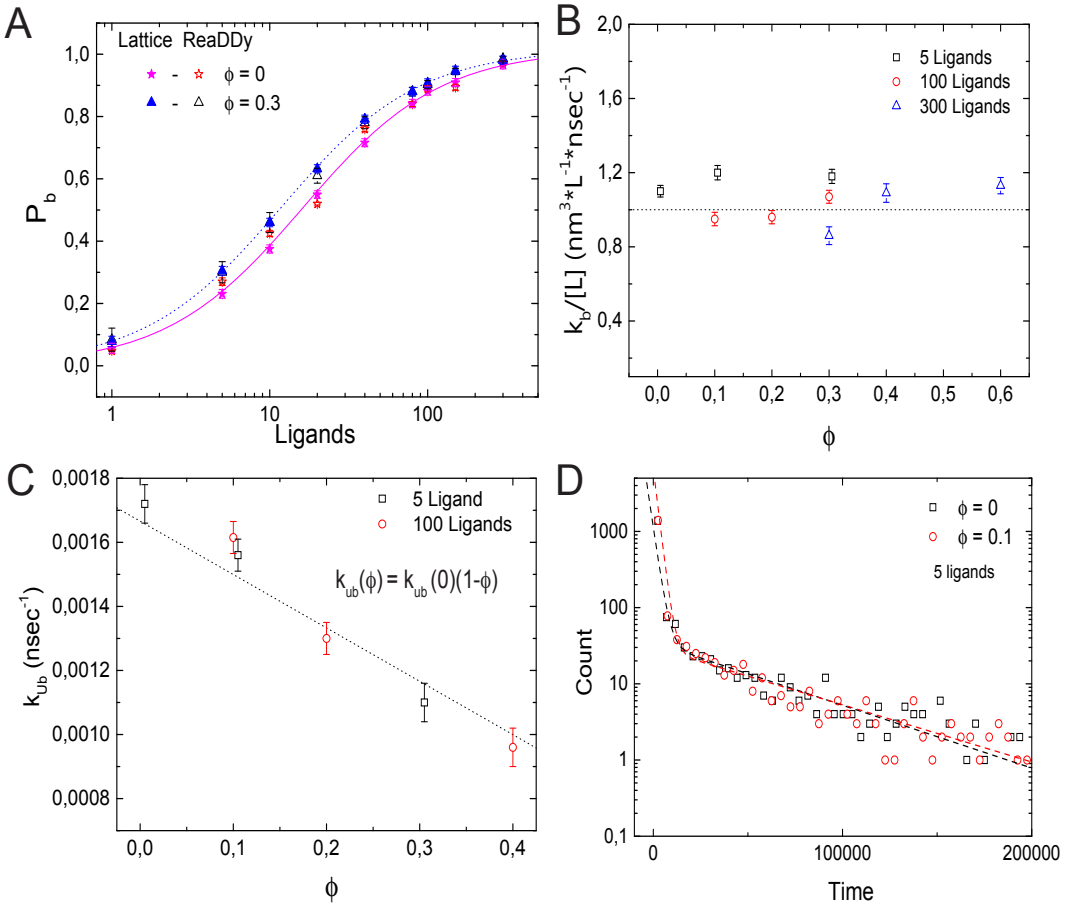

Figure 2.4 – Effects of molecular crowding on binding equilibria simulated with ReaDDy. (A) Probability for the target (receptor) to be occupied P_b as a function of the number of ligands, for two levels of volume occupation $\phi = 0$ and 0.3. Crowding shifts the probabilities towards lower number ligands. Similar behavior is observed for the lattice implementation and for ReaDDy. (B) Binding rate k_b stays constant as the levels of volume occupation increase. (C) Unbinding rate decays as ϕ increases, following the relation $k_{ub} = k_{ub}(0)(1 - \phi)$. (D) Histograms of 1700 binding events for two different values of volume occupation $\phi = 0$ and 0.3. A double exponential decay is observed, indicating the presence of a slow and a fast component of the dynamics, as in the lattice implementation. Simulation parameters: $V = 49\text{x}49\text{x}49$ nm^3, $k_{ub}(0) = 1.6\text{x}10^6$ sec^{-1}, $r = 3$ nm, $D = 166.6$ μm^2 sec^{-1} and $T = 20$°C. The data from lattice simulations in (A) are the same as in Fig. 2.2A.

the lattice model, the unbinding rate follows $k_{ub} = k_{ub}(0)(1 - \phi)$. We can thus conclude that the effects of crowding on P_b are due to the slower dissociation rate. Next, in Fig. 2.4D we plot the histograms of 1700 binding events for the cases $\phi = 0$ and 0.1. As in the lattice implementation, the histograms follow a double exponential distribution ($R^2 = 0.996$ and 0.998 for $\phi = 0$ and 0.1, respectively). Thus, simulations with ReaDDy also show a slow and a fast component of the binding kinetics. We note that although for both implementations the simulation time increases quadratically with particle number, the simulation times in ReaDDy are about 2 orders of magnitude larger than in our implementation (data not shown). This is primarily due to the computationally expensive evaluation of the steric potential in ReaDDy.

2.4 Effect of crowding on diffusion-limited reactions

We have seen above that crowders do not affect the rate of formation of the receptor-ligand complex, because they suppress binding events involving large-distance diffusion of the ligand to the receptor, but also enhance (re-)binding for ligands close to the receptor and the two effect compensate each other. We now consider the reaction between an enzyme (replacing the receptor) and its substrate (replacing the ligand). Specifically, we consider a reaction that is diffusion-limited and very efficient, converting every incoming substrate into the corresponding product P, i.e. we consider the irreversible reaction $L + R \to LR \to P + R$. In contrast to the receptor-ligand binding considered above, in this case, immediate rebinding of the substrate is not possible, because the released molecule is the product rather than the substrate. Thus, one can expect that the balance between the different effects of the crowder is perturbed and different behavior can be expected in this case. Clearly, the reaction we consider here is a limiting case; a more general and more realistic scenario, namely Michaelis-Menten kinetics, that allows both the unbinding of the substrate and release of the product will be considered below.

2.4.1 Lattice simulations

We study the enzyme-substrate reaction under steady-state condition with constant substrate and product concentrations. To implement this case, we use the same simulations as above with only a difference in the unbinding process. Instead of unbinding of the ligand, we now simulate a reaction that transforms the substrate into the product. In addition, we need to implement an additional reaction that keeps the substrate and product concentrations constant. In our lattice simulation, these two events are implemented in one single step: When a substrate is bound to the enzyme (i.e., occupies the target site), the reaction occurs with rate k_r and simultaneously the product is released and removed from the simulation and a new substrate molecule is introduced at a random position. Thus, we keep the

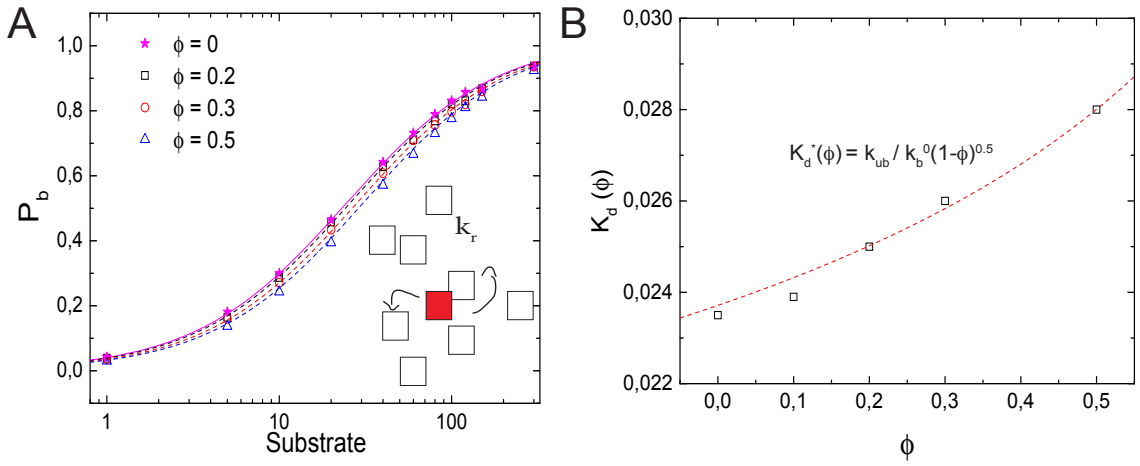

Figure 2.5 – Effects of molecular crowding on a diffusion-limited reaction. (A) Probability P_b for the target (enzyme) to be occupied P_b as a function of the number of substrates, for $\phi = 0, 0.2, 0.3$ and 0.5. Molecular crowding shifts P_b towards larger number of ligand. Thus, more ligands are needed to saturate the receptor when ϕ increases. (Inset) Representation of the reaction model, in which a product P is synthesized with a rate k_r. (B) Dissociation constant $K_d^*(\phi)$ (obtained as the substrate concentration for which $P_b = 0.5$) as a function of ϕ, which increases as a function of ϕ. The line is obtained from Eq. 2.7 with $\kappa = 0.5$, as obtained from the fit in Fig. 2.6B. Simulation parameters as in Fig. refBiRe2, and $k_r = 1/60\ \tau^{-1}$.

product concentration zero and the substrate concentration at its initial value. Effectively, the reaction is described as a unbinding process to a random position in space rather than a neighbor site of the enzyme. Clearly, this dynamics does not satisfy detailed balance; energy input is required to keep the concentrations and thus the chemical potential constant.

In Fig. 2.5A, we plot the probability that the target site, the enzyme, is occupied as a function of the number of substrates in the box. In contrast to the ligand-receptor binding reaction considered above, increased crowder numbers now shift the function to the right, that is, towards larger ligand numbers, so that the occupation of the enzyme (as well as the overall reaction rate) is reduced by the crowders. We note that the effect is rather modest, but nevertheless it is important that the effect is the opposite of what we observed above. Although the process does not correspond to a binding equilibrium, the data are well-described by Eq. 2.6, with an effective dissociation constant $K_d^*(\phi) = k_r/k_b$, where k_r is the reaction rate and k_b the diffusion-limited binding rate. The effective dissociation constant (determined as the substrate concentration for which $P_b = 0.5$) is plotted in Fig. 2.5B as a function of the volume fraction of crowders. It increases with increasing volume fraction and can be described by

$$K_d^*(\phi) = \frac{k_r}{k_b} = \frac{k_r}{k_b^0(1-\phi)^\kappa} \tag{2.7}$$

with $\kappa \simeq 0.5$ (obtained from the fit to the binding rate below).

This functional form is not entirely surprising: Our reaction is diffusion-limited, thus the binding rate is proportional to the diffusion coefficient ($k_b = 4\pi\sigma D$, where $\sigma \simeq l$ is the reac-

tion radius of the interacting particles). Experiments for different concentrations and species of crowding agents have shown that the change in the diffusion constant of different proteins due to crowding can be described by the phenomenological expression $D(\phi)/D(\phi = 0) = (1 - \phi)^{\kappa}$ (14; 22). In the presence of different crowding agents, diffusion measurements for the protein carbon monoxide hemoglobin have been fitted with a value of $\kappa = 0.36$ (14; 22).

That our result for the effective dissociation constant shows the same dependence as experimentally observed for diffusion indicates that, in contrast to the binding reaction studied above, here the entire effect of crowding is via the (diffusion-limited) binding rate. This is indeed the case, as one can see in Fig. 2.6A and 2.6B, where we plot the reaction rate (which serves as an effective unbinding rate) and the binding rate individually as functions of the volume fraction, respectively. The bare reaction rate is unaffected by crowding, while the binding rate decreases with increasing volume fraction, with the same functional dependence as the effective dissociation constant,

$$k_b(\phi) = k_b^0(1 - \phi)^{\kappa}. \tag{2.8}$$

This dependence was used for the fit to determine $\kappa \simeq 0.5$ (with correlation coefficient $R^2 = 0.98$ and 0.96 for 10 and 100 substrates, respectively). As a side remark, we note that the results for the binding rate of this reaction can be used to determine the effect of crowding on the diffusion coefficient, which would be more difficult to obtain from the mean square displacement due to the effect of confinement in the finite simulation volume.

Next, we consider the histograms of the ligand binding times, which are shown in Fig. 2.6C. In contrast to the binding reaction considered above (Fig. 2.3D), the simulation data for this reaction are well-described by a single exponential (correlation coefficient of fit $R^2 = 0.998, 0.995$ and 0.998 for $\phi = 0, 0.3$ and 0.5, respectively). When ϕ is increased, the characteristic time increases, that is, the ligand-receptor encounter needs more time to take place. This result corresponds to the slow component of the binding scenario discussed above. In fact the two results show good quantitative agreement. We therefore conclude that our description of a diffusion-limited reaction completely uncouples the effect of molecular crowding on diffusion from the effect on binding equilibria.

2.4.2 Off-lattice simulations

Next, we implement this reaction in ReaDDy. We note that the reaction described on the lattice is idealized in the sense that removal of the product and re-introduction of substrate to keep the substrate concentration constant occur simultaneously at different positions on the lattice. For the off-lattice case, this idealization is less straightforward and it is easier to uncouple these processes. Thus, we include an additional reaction in these simulations: A product can be converted back into a substrate with rate k_c. Indeed such processes are real-

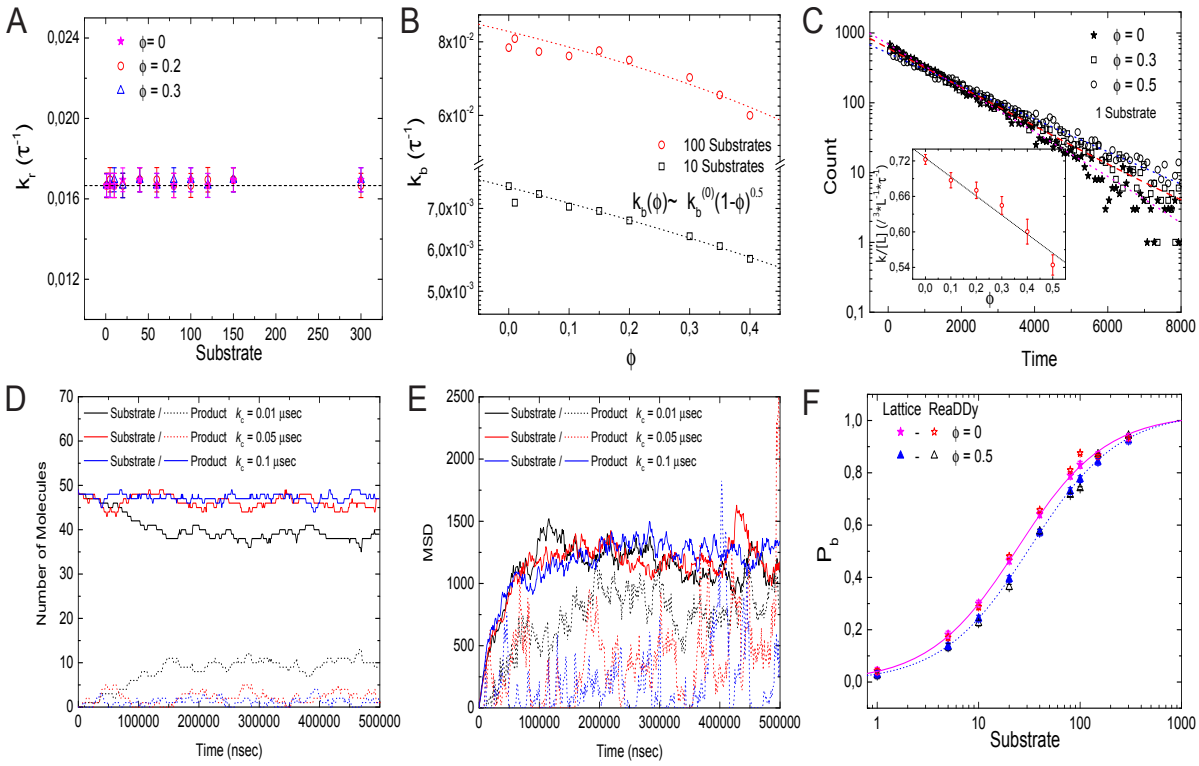

Figure 2.6 – Kinetics of the diffusion-limited reaction. (A) Reaction rate k_r as a function of the number of substrate molecules for different levels of volume occupation ϕ. k_r stays constant and has a value of $k_r = 1/60\tau^{-1}$ for all substrate numbers and levels of ϕ. (B) Binding rate k_b decays as ϕ increases. The simulation data is fitted with Eq. 2.8, resulting in $\kappa \simeq 0.5$ (lines). (C) Binding time histograms fitted with a single-exponential. (Inset) The characteristic time increases as the levels of volume occupation increase. (D, E) Trajectories of the implementation of the diffusion-limited reaction with ReaDDy. Number of particles (D) and mean square displacement (E) from the target as functions of time for different charging rates. The solid lines represent the substrate particles (ligands) that can bind and react, the dashed lines represent the product that needs to be recharged with rate k_c before it can bind again. (F) Comparison of the binding probability P_b as a function of the number of ligands at different crowding levels for ReaDDy simulations and the lattice model (data from Fig. 2.5A). Simulation parameters (ReaDDy): $V = 49\text{x}49\text{x}49$ nm^3, $k_r = 1.6\text{x}10^6$ sec^{-1}, $r = 3$ nm and $T = 20$°C.

ized for example in translation, where tRNAs leave the ribosome uncharged, i.e. not carrying an amino acid and get recharged by tRNA synthetases. The product constitutes an additional molecular species which we take to diffuse with the same diffusion coefficient as the substrate, but not to bind to the enzyme/target.

To compare the ReaDDy simulations with our lattice model, two difficulties must be solved: In the first place, we want to keep the substrate (charged ligand) concentration as unaffected as possible. On the other hand, the product or uncharged ligand must diffuse far from the receptor and be charged at a random position within the simulation box. These two requirements are antagonistic, meaning that on average, fast recharging implies short diffusion, and long diffusion implies slow recharging. Thus, we run simulations for different values of the recharging rate k_c in order to obtain an adequate value to simulate our ideal reaction.

Fig. 2.6D shows simulation time courses for the number of molecules with 50 charged and no uncharged ligands as initial condition. For a small recharging rate, $k_c = 0.01\ \mu\text{sec}^{-1}$, the charged ligand number decreases to about 40 molecules. For a fast charging rate $k_c = 0.1\ \mu\text{sec}^{-1}$, the charged ligand number stays almost constant at 50 molecules with slight fluctuations. However, the mean square displacement of the substrate and product particles from the origin of the simulation box, which we use as a measure of the distance of the target, shows that for the high recharging rate, the products (uncharged ligand) are on average closer to the target than the substrate/charged ligand. Even for the low recharging rate, a small difference in mean square displacement can still be seen. To balance the two requirements, we therefore used an intermediate charging rate of $k_c = 0.05\ \mu\text{sec}^{-1}$. For this value, the number of charged ligands does not change drastically (Fig. 2.6D), and the uncharged ligand diffuses far from the ligand before it gets recharged, see Fig. 2.6E.

Using this intermediate value of the recharging rate (and all other simulation parameters as above), we determine the target occupation P_b as a function of the number of ligands for different levels of crowding (Fig. 2.6F). As the volume occupation increases, P_b is shifted to the right, towards larger ligand concentrations, in quantitative agreement with the lattice simulations.

2.5 Effects of crowder size

So far, we have only considered crowders with the same size as the diffusing ligand or substrate molecules. However, effects of molecular crowding on biochemical reactions are known to be dependent on the reactants' size and geometry (8). Specifically, large particles experience entropic attraction forces in the presence of smaller crowders. Such forces are known as depletion forces and have been studied extensively, both theoretically and experimentally (131; 132).

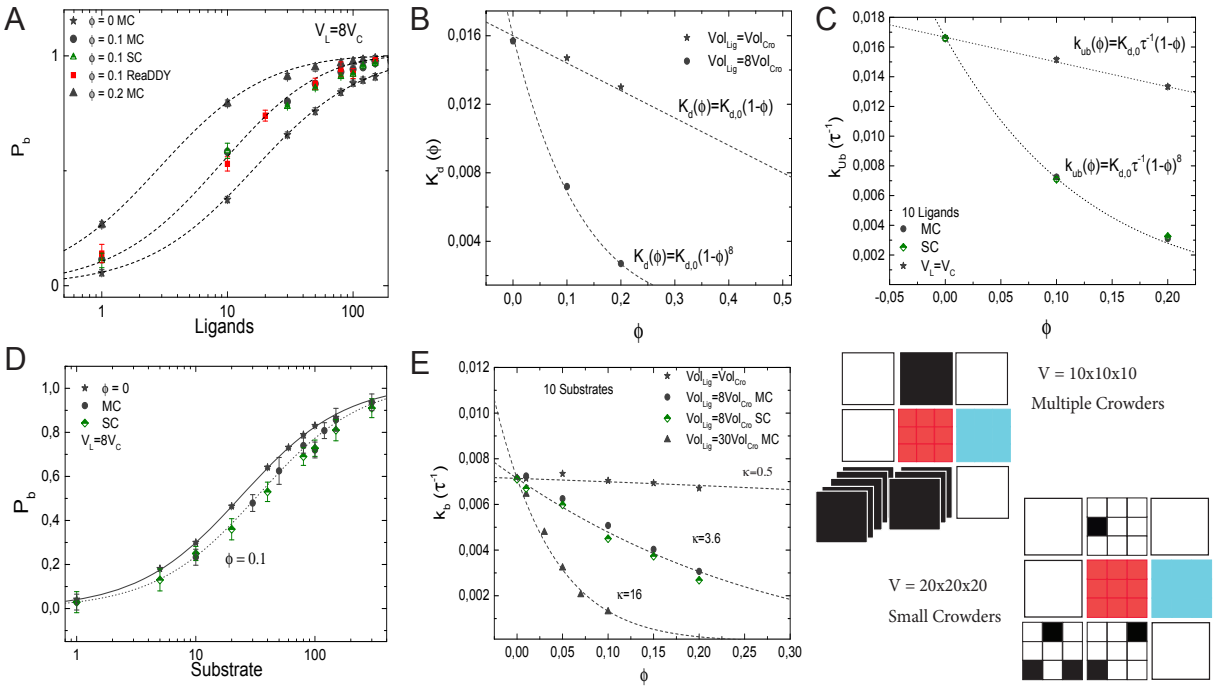

Figure 2.7 – Effects of crowder size on the binding of a ligand to a receptor (A-C) and on the diffusion-limited enzymatic reaction (D, E). In all panels, grey-filled symbols represent the implementation in which multiple crowders can occupy the same lattice site (MC), half-filled symbols (green and pink) represent the implementation of the small crowders (SC) and the red symbols show results from ReaDDy. In general, good agreement between the methods is observed. (A) P_b of the simple binding reaction as a function of the number of ligands for different crowder sizes and volume occupation fractions ϕ. The lines are from Eq. 2.9 with no free parameters. (B) Dissociation constant as a function of ϕ for two different crowder sizes. Smaller crowders have a stronger effect on the dissociation constant than crowders with the same size as the ligands. (C) Unbinding rate as a function of crowder volume fraction. (D) P_b for a diffusion-limited reaction at different crowder volumes. (E) Corresponding binding rate. The lines show fits with Eq. 2.8 ($R^2 = 0.96, 0.995$ and 0.994 in descending order). The simulation parameters are as in Fig. 2.2 and Fig. 2.5.

Here we are interested in the case of crowders smaller than the ligand/substrate. In the lattice simulations we thus consider ligands occupying more than one lattice site, specifically cubic particles occupying 8 sites. We implemented this situation in two ways, explicitly with crowders occupying a single site and ligands occupying $n = 8$ sites (small crowders model, SC) and in an approximate way, where the different sizes are taken into account implicitly, by allowing up to $n = 8$ crowders to occupy the same site, while the presence of a single crowder already excludes a ligand from that site (multiple crowders model, MC). In the latter, n is the maximal number of crowders allowed to occupy the same lattice site (which can have any integer value, while in the first model only sizes $n = k^3$ with integer k are possible). We note that in the first implementation, the lattice constant corresponds to the size of the crowder, while in the second it corresponds to the size of the ligand. Thus, to compare the two implementations, one needs to account for the different unit lengths as well as the corresponding different unit times, which are defined via the diffusion over the lattice spacing.

The effect of the crowders on the binding equilibrium can be calculated in the same way as above, by counting the number of configurations S_{ub} and S_{b} by distributing L or $L-1$ ligands and C crowders on the lattice (5). This leads to the probability that the receptor is occupied as given by

$$P_b = \frac{1}{1 + \frac{K_d(\phi)}{[L]}} \qquad \text{with} \qquad K_d(\phi) = K_{d,0}(1-\phi)^n. \tag{2.9}$$

Thus, small crowders have a stronger effect on binding than crowders of the same size as the ligands. This results is confirmed by the simulations. Fig. 2.7A shows P_b as a function of the number of ligands for different levels of molecular crowding. Here, grey-filled symbols show data from simulations where multiple crowders can occupy one lattice site, half-filled colored symbols (pink and green) represent simulations with small crowder particles on the lattice and the red-filled symbols show results from ReaDDy simulations. For the latter, crowders have a radius $r = 1.5$ nm to obtain a volume 8 times smaller than the ligands. Symbols show averages of our simulations after running over 3000 binding events. Half-saturation of the receptor at $\phi = 0.2$, is reached at $\simeq 3$ ligands, whereas for the case where crowders and ligands have the same size, i.e $n = 1$, half-saturation occurs at $\simeq 13$ ligands. Lines represent Eq. 2.9 and show good agreement with the simulations. The effective dissociation constant is plotted, for the multiple crowder approximation, as a function of the volume fraction ϕ in Fig. 2.7B. Comparison of the data for crowders with the same size as the ligand and with an 8-fold smaller volume, shows that for the same occupied volume fraction, the small crowders shift the dissociation constant more strongly toward the bound state than the larger crowders. Thus, only for crowders of the same size as the ligand, the effect of crowding can be identified with a reduction of the available volume by the volume occupied by the crowders. For the smaller crowders, the available volume is reduced by a larger amount. This observation can be explained by the fact that the volume from which large particles are excluded is determined

by the spatial arrangement of small particles. This additional volume exclusion effect can also be interpreted as due to an additional attractive force between the large particles (here the ligand and the target). The effect of crowder size is also known from scaled particle theory, however in that theory non-linear ϕ-dependencies are already present for crowders of the same size as the ligands (8).

Next, we plot the unbinding rate as a function of the volume fraction for smaller crowders, see Fig. 2.7C. Independent of the crowder size, the effect of crowding on the dissociation constant is entirely mediated by the unbinding rate. Results are in good agreement with the expression for the unbinding rate $k_{ub} = K_{d,0}\tau^{-1}(1-\phi)^n$, with $n = 1$ and 8 for crowders with the same size as the ligands, and crowders 8-fold smaller than the ligands, respectively. No strong difference between the two different implementations is observed.

Next, we study the effect of small crowders on ligand diffusion using again the diffusion-limited reaction in which the target converts a substrate (the ligand) into a product (which is instantaneously replaced by a new substrate inserted into the simulation box at a random position). As for the crowders with the same size as the ligand, binding to the target is weaker in the presence of the crowders and the dissociation constant increases with increasing volume fraction due to a reduction of the binding rate (the effective unbinding rate, given by the reaction rate, remains constant by construction of the reaction model). Fig. 2.7D shows P_b as a function of L for two different volume fractions of crowders, which are again 8-fold smaller than ligands. In Fig. 2.7E, we plot the binding rate k_b as a function of volume fraction for a fixed number of ligands, $L = 10$, and three different crowder sizes (grey and green symbols). When crowders and ligands have the same size $n = 1$, the decrease in the binding rate is modest. Reducing the sizes of the crowders increases the effect of crowding. As above, the dependence of the binding rate on the volume fraction can be described by $k_b = k_b^0(1-\phi)^\kappa$, with $\kappa \simeq 0.5$, 3.6 and 16 for crowders with the same size as the ligands, 8-fold smaller than the ligands and 30-fold smaller than the ligands, respectively. For the case of crowders being 8-fold smaller than the ligands, where we tested both implementations of smaller crowders, no strong difference is observed between the two different implementations.

2.6 Enzymatic reaction

Finally, to consider the combined effect of crowding on diffusion, binding and reactions, we simulate an enzymatic reaction with Michaelis-Menten kinetics (133), i.e. $L + R \Leftrightarrow LR \rightarrow P + R$. This reaction is implemented in our lattice model by combining the two reactions studied above: The substrate binds to the target (the enzyme) as before, but then two events may occur. The substrate may unbind from the target with rate k_{ub} as in the binding reaction studied in section III. Alternatively, the reaction can take place and a product is released with rate k_r. In this case, we again remove the product from the simulation box and re-insert a

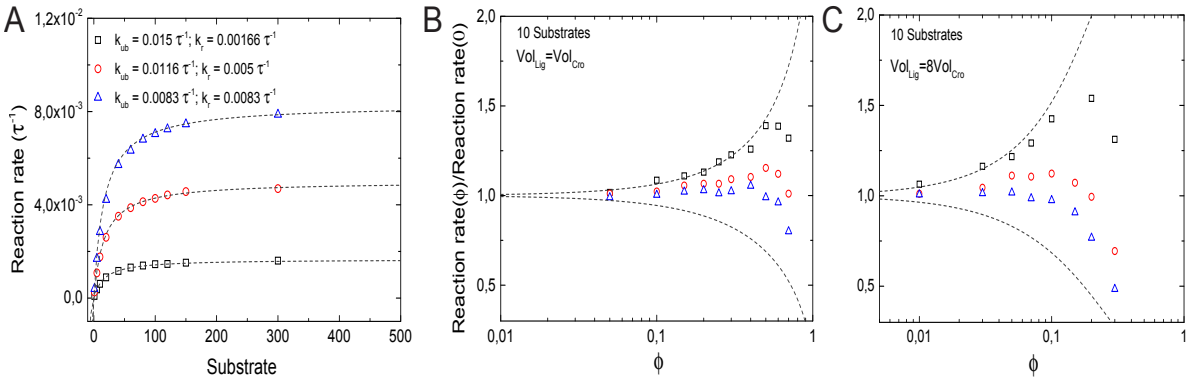

Figure 2.8 – Effects of molecular crowding on a reaction with Michaelis-Menten kinetics. (A) Overall product synthesis rate in absence of crowding ($\phi = 0$) as a function of the number of substrates for different set of parameters k_{ub} and k_r (in units of the inverse simulation time steps τ^{-1}, all other parameters are as in Fig. 2.2). (B) Ratio between the overall synthesis rate in the presence of crowders and the overall synthesis product rate without crowding as a function of crowder volume fraction ϕ. Non-monotonic behavior is obtained for all sets of parameters. For small values of ϕ, binding equilibrium effects are dominant, shifting the ratio to values > 1; for large values of ϕ, the slow diffusion dominates, as the reactions become diffusion-limited and the ratio decreases again. The maximal reaction rate is found at an intermediate value of ϕ. (C) The same as in (B), but for crowders 8-fold smaller than substrates. The qualitative behavior is the same as in (B), but the presence of smaller crowders shifts the maximums towards lower values of ϕ. The dashed lines in (B, C) represent the theoretical limits of $k_{ub} \gg k_r$ (top) and $k_{ub} \ll k_r$ (bottom).

substrate molecule at a random position. In the absence of crowders, the synthesis of product proceeds with a rate

$$\frac{dP}{dt} = \frac{k_r[R][L]}{K_M + [L]}, \tag{2.10}$$

with the Michaelis constant $K_M = (k_r + k_{ub})/k_b$. In the two limiting cases where either k_r or $k_{ub} = 0$ are negligibly small, the reaction reduces to the two reactions considered above. In the following, we will modulate these two parameters such that $k_{ub} + k_r$ is kept constant ($= 1/60\tau^{-1}$). Thus, in the absence of crowding, the reactions have the same Michaelis constant, but differ in their maximal reaction rate, see Fig. 2.8A. By contrast, when crowders interfere in the reaction, the maximal reaction rate may be limited by different physical processes. If unbinding is negligible, the reaction is diffusion-limited, and thus hindered by crowding. The Michaelis constant is increased by crowders, $K_M \approx k_r/k_b \times (1-\phi)^{-\kappa}$. As we have seen, this increase can be attributed to a reduced binding rate, and the value of κ is strongly dependent on crowder size. In the other limit, the reaction is limiting and the binding/unbinding process has enough time to reach equilibrium before a reaction takes place. In that case, the Michaelis constant is decreased by the crowders, $K_M \approx k_{ub} \times (1-\phi)^n/k_b$ due to a reduced unbinding rate. For a given concentration of substrate, the rate of product formation is reduced in the first case, but increased in the second. These opposing limiting behaviors are plotted as dashed lines in Figs. 2.8B and 2.8C.

The latter competitive effects lead to a non-monotonic behavior of the reaction rate for parameters that lie between the limit cases, as shown by the symbols in Fig. 2.8B and 2.8C. In these cases, the reaction rate is increased by crowding at low volume fractions of crowders, but decreases for large volume fractions. This observation can be explained as a transition between the reaction-limited and the diffusion-limited case, as increasing volume fractions decrease the unbinding rate. Thus, the volume fraction for which the reaction rate is maximal can be estimated by $k_r = k_{\mathrm{ub}}(1-\phi)^n$, which leads to a diffusion-limitation for $\phi \gtrsim 1 - (k_r/k_{\mathrm{ub}})$ for $n = 1$. This estimate also indicates that for smaller crowders (larger n), the maximal reaction rate occurs for smaller volume fractions. This expectation is confirmed by simulations, shown in Fig. 2.8C. Here decreasing reaction rates are seen for volume fractions below the typical intracellular crowding level of 0.3.

Experimentally, both increases and decreases of enzyme activity with increasing levels of crowding have been seen (4). However, often these observations do not reflect simply the shift in equilibrium binding, in particular, when substrates are small compare to the crowders. Rather, crowding can also affect the activity of an enzyme by modulating its conformation or by inducing oligomerization (20; 43). However, there is evidence for decreased activity due to diffusion-limitation in several cases (20; 134).

2.7 Discussion: Crowding effects on gene expression

In the preceding sections we have discussed generic effects of molecular crowding on two elementary types of reactions, simple binding/unbinding and an enzymatic reaction converting a substrate into a product. These two reaction paradigms can be used to describe many different processes in cells, including some that are not enzymatic in a strict sense. An example for the latter is binding of RNA polymerase to a promoter and initiation of transcription, which can be described (within a minimal mathematical representation) by Michaelis-Menten kinetics with the promoter taking the role of the enzyme (135; 136). In the following, we will thus discuss some applications of these two elementary reactions to processes in gene expression.

2.7.1 Transcription factors

Most efforts to study the impact of molecular crowding on gene expression have been devoted to the binding of transcription factors to their binding sites on the chromosome (13; 14; 137). For most transcription factors, this is a typical case of equilibrium binding and indeed most models for gene regulation are based on the assumption of a binding equilibrium for transcription factors (138; 139). As discussed above as well as in a large body of previous work, one generically expects such binding to be strengthened by crowding. Interestingly, this holds

both for specific binding to the functional binding sites and for (sequence-independent) non-specific binding. Since the molecules involved are the same, the relative strength of specific and non-specific binding should not be affected. Thus, the increase in binding is mainly at the cost of the cytoplasmic fraction of the transcription factors, in agreement with the observation that transcription factors spends most of the time bound to DNA (32; 140).

The strengthening of non-specific binding by crowding should also have an interesting consequence for the dynamics. Transcription factors diffuse in the cell by a combination of three-dimensional cytoplasmic diffusion and one-dimensional diffusion (sliding) along DNA while non-specifically bound (32). The one-dimensional diffusion coefficient is typically considerably smaller than the diffusion coefficient for three-dimensional diffusion. Thus, unless the transcription factor is bound to DNA in close proximity to a specific binding site, where sliding plays an important role, non-specific binding to DNA can mostly be interpreted as inhibiting cytoplasmic diffusion. Thus the effective diffusion coefficient can be approximated as $D_{\text{eff}} = D(1 - P_{\text{b,ns}})$, where D is the cytoplasmic diffusion coefficient. Assuming that the crowders in the cell are mostly proteins and thus similar in size to the transcription factor of interest, one should expect crowding to have a relatively mild effect on the transcription factor's diffusion in the cytoplasm with $D(\phi) \simeq D(1 - \phi)$. However, since non-specific binding is strengthened by crowding, the inhibition of cytoplasmic diffusion is also enhanced and cytoplasmic diffusion is interrupted by pauses on the DNA that get longer and longer with increasing volume fraction of the crowders. Recent Brownian Dynamics simulations (28) indeed showed an increase in the fraction of time spend bound to DNA, however the overall effect on search times, the time required to find a binding site on the DNA, was found to be only weakly affected by crowding, as the opposing effects of crowding seem to keep each other in balance.

2.7.2 Transcription

As mentioned above, the initiation of transcription can be described by Michaelis-Menten kinetics with RNA polymerase reversibly binding to the promoter and irreversibly starting to elongate an RNA chain. Thus, the promoter formally takes the role of the enzyme and converts free RNA polymerases into transcribing RNA polymerases. Contrary to many transcription factors, RNA polymerase is a relatively big protein with a molecular weight of ≈ 400 kDa (141). Thus, its size is larger than that of the typical crowder and crowding effects can be expected to be more pronounced. However, to address the effect of crowding on the initiation of transcription, we first need to estimate whether this reaction is diffusion-limited. For a large protein complex such as RNA polymerase, the cytoplasmic diffusion coefficient is approximately 1 $\mu\text{m}^2/\text{s}$ (126; 142), which is reduced due to non-specific binding to approximately 0.2 $\mu\text{m}^2/\text{s}$ (142; 143). Thus, the diffusion-limited binding rate to a promoter can be estimated to be about 0.1 $\mu\text{M}^{-1}\text{s}^{-1}$. This value could be increased due to sliding along the

DNA, which effectively increases the size of the target to be reached by cytoplasmic diffusion. However, the importance of sliding is unclear, recent studies have questioned it plays an important role at all and it will clearly be limited by the presence of other DNA-bound proteins including the transcription factors bound near a promoter. Since typical transcription rates are of the order of a few per minute, however, for most cases, with a concentration of RNA polymerases of 5-10 μM, transcription should not be limited by diffusion. An exception might be the transcription of ribosomal RNA, which exhibits much larger transcription rates, up to 80 per minute. Notably, the cellular RNA polymerase pool is quite large, exceeding numbers needed for transcription. One can speculate that a smaller pool would make transcription initiation diffusion-limited at least for highly transcribed genes such as those encoding ribosomal RNA and thus not be sufficient for the high transcription rates required on these genes.

However, for most genes, the initiation of transcription should not be limited by diffusion of RNA polymerase and thus crowding can be expected to enhance polymerase binding to the promoter and thus transcription.

2.7.3 Translation

Similar to the case of RNA polymerase, one can argue that the initiation of translation could be limited by diffusion of the ribosome, which is an even bigger molecular machine than RNA polymerase; however it does not exhibit non-specific binding to DNA. Contrary to RNA polymerase however, the pool of free ribosomes appears to be rather small (144). To a first approximation, ribosomes are translating all the time, an observation that can be interpreted as efficient use to maximize the return of an expensive investment (145; 146). Thus, initiation of translation on any specific mRNA could well be limited by the diffusion of ribosomes. However, different mRNAs compete for ribosomes and such a limitation would not result in inefficient use of ribosomes but rather in the translation of a different mRNA. Thus, the cell's objective here may not be affected by a diffusion limitation.

The elongation process is also quite different for translation compared to transcription. Elongation of the growing polypeptide chain requires that the next amino acid is delivered to the ribosome by a ternary complex containing a tRNA charged with the amino acid and a GTP-activated elongation factor Tu. This complex is again a large molecular complex with a small diffusion coefficient and a molecular size exceeding the size of typical crowders. Assuming a concentration of a few μM for typical ternary complex species (147), binding to the ribosome is expected to occur with rate $\sim 10\ \mathrm{s}^{-1}$ (127). Thus, peptide chain elongation may proceed rather close to the diffusion limit and the large concentration of ternary complexes in cells (EF-Tu is the most abundant protein in *E. coli* cells (128)) is likely required to avoid such limitation to ensure efficient cellular use of ribosomes. Indeed, from a proteome partitioning point of view, the optimal solution would be to set the Michaelis constant of translation

elongation as low as possible. Thus, the actual value must be set by some limitation such as a diffusion limitation, which results in a lower limit for the Michaelis constant (127). Thus any increase in the level of crowding should slow down translation and, via to the close link between protein synthesis and cell growth, have a negative effect on cell growth. We note, however, that such a limitation could be circumvented by local ternary complex pools. Indeed it has been suggested that tRNAs are recharged while associated with the ribosome (148). The dynamics is then similar to a high recharging rate in the recharging process in our ReADDy simulations (Figure 2.6) and crowding would not have the expected negative effect. However, definitive proof for such local recharging is still lacking.

2.8 Concluding remarks

In this paper, we used a simple computational approach to discuss the effects of molecular crowding on several simple enzymatic reactions, specifically for relatively big molecular substrates. We used a combination of lattice and off-lattice simulations to revisit the two elementary consequences of molecular crowding, namely enhanced binding and reduced diffusion. The balance between these two effects can be subtle as indicated by the example of the binding rate, which remains unaffected due to an increase in rapid rebinding events and, at the same time, a decrease in binding event involving diffusive arrival of ligands. The lattice model provides a rather intuitive picture of these situations (as well as a very efficient computational implementation).

In addition, we have discussed applications of these effects to steps in gene expression, such as transcription factor-DNA binding, promoter finding by RNA polymerase and translation elongation. Our estimates show that some of these processes may come close to the diffusion limit and that such diffusion limitation may be physiologically important, specifically for translation elongation.

The methods we used here can be applied more generally and there are a number of possible extensions to this work, such as addressing the effects of spatial clustering of targets (binding sites/enzymes) and the built-up of local concentrations.

Chapter 3

Facilitated diffusion in the presence of obstacles on the DNA

Abstract

Biological functions of DNA depend on the sequence-specific binding of DNA-binding proteins to their corresponding binding sites. Binding of these proteins to their binding sites occurs through a facilitated diffusion process that combines three-dimensional diffusion in the cytoplasm with one-dimensional diffusion (sliding) along the DNA. In this work, we use a lattice model of facilitated diffusion to study how the dynamics of binding of a protein to a specific site (e.g., binding of an RNA polymerase to a promoter or of a transcription factor to its operator site) is affected by the presence of other proteins bound to DNA that act as 'obstacles' in the sliding process. Different types of these obstacles with different dynamics are implemented. While all types impair facilitated diffusion, the extent of the hindrance depends on the type of obstacle. As a consequence of hindrance by obstacles, more excursions into the cytoplasm are required for optimal target binding compared to the case without obstacles.

3.1 Introduction

Processing of the genetic information is to a very large extent dependent on the sequence-specific binding of proteins to DNA. Examples include transcription factors binding to specific operator sites and RNA polymerases binding to promoters (26). Sequence-specificity is typically not perfect, and binding motifs therefore are typically 'fuzzy', with a range of different sequences displaying comparable affinities for the protein and an even large range showing weak affinity (27). In addition, binding of DNA-binding proteins (DBP) to DNA usually has an electrostatic component that is independent of sequence, i.e. non-specific. For that reason both the equilibrium binding pattern and the binding kinetics of DBPs can be quite complex (149). Nevertheless, DBPs find their functional binding sites (of size $\sim 10 - 30$ bp (26; 66))

with remarkable efficiency despite the large number of non-specific binding sites ($\sim 10^6$ in the genome of *E. coli*) that compete for binding with the functional site. In a seminal work, Riggs *et al.* measured the *in vitro* rate at which the *lac* repressor finds its promoter, and they found that under certain conditions, the *lac* repressor binds to its target at faster rates than the simple three dimensional diffusion limit (68). To explain this phenomena, Berg, Winter and von Hippel proposed the so called facilitated diffusion model (69; 70), in which such high binding rates are achieved if the DBP undergoes a combination of three-dimensional (3D) excursions in the bulk solution, together with one-dimensional (1D) sliding on the DNA. This theoretical model has been strongly supported by experimental techniques that directly showed the number of basepairs scanned via 1D sliding (71; 72; 73), and the average fraction of time the DBP remains bound to the DNA before unbinding (32; 74).

Since it was proposed, facilitated diffusion has been the subject of much theoretical (76; 79; 80; 81; 82; 150), experimental (66; 83; 84; 85; 86; 151; 152) and computational (28; 75; 78) efforts. Interest in facilitated diffusion has been renewed by the direct observation of facilitated diffusion in bacterial cells using single-molecule techniques (32; 74). One key feature that is different in cells compared to *in vitro* is that cytoplasm is not a dilute solution, but a rather crowded environment that can be occupied up to 40% by macromolecules (1; 87). The presence of these macromolecules (crowders) inside the cells has effects on diffusion (22; 23; 24), enzymatic reactions (4; 20; 21; 25), protein folding (9; 10; 11; 19) and gene expression (13; 14; 15). In addition, the DNA itself is also covered with proteins that bind to the DNA in order to perform functions such as transcription, DNA repair and gene regulation (26; 29; 30; 31). Moreover, the DNA itself is spatially organized and compacted by histones in eukaryotes and nucloid-associated proteins in bacteria (153; 154). Thus, the DNA is also highly occupied ($\sim 30\%$) by DBPs that affect the facilitated diffusion process. So far, only few of the theoretical studies have included the impact of these macromolecules on the promoter finding dynamics (33; 88; 89; 90; 91).

In this paper we address the effects of obstacles bound to the DNA on facilitated diffusion, i.e. on the promoter search of RNA polymerase or the search of a transcription factor for its functional binding site. We use a lattice model to study facilitated diffusion, first in the free case without crowders and then in the presence of obstacles on the DNA. Specifically, we consider three different types of obstacles characterized by different dynamics on the DNA. We show that, for all types, facilitated diffusion is impaired when the DNA occupation fraction increases, but the strength of the effect depends on the type of obstacles. The lattice model we use aims to create a clear conceptual understanding of facilitated diffusion rather than provide a detailed computational description of the process. The method is, however, applicable more generally.

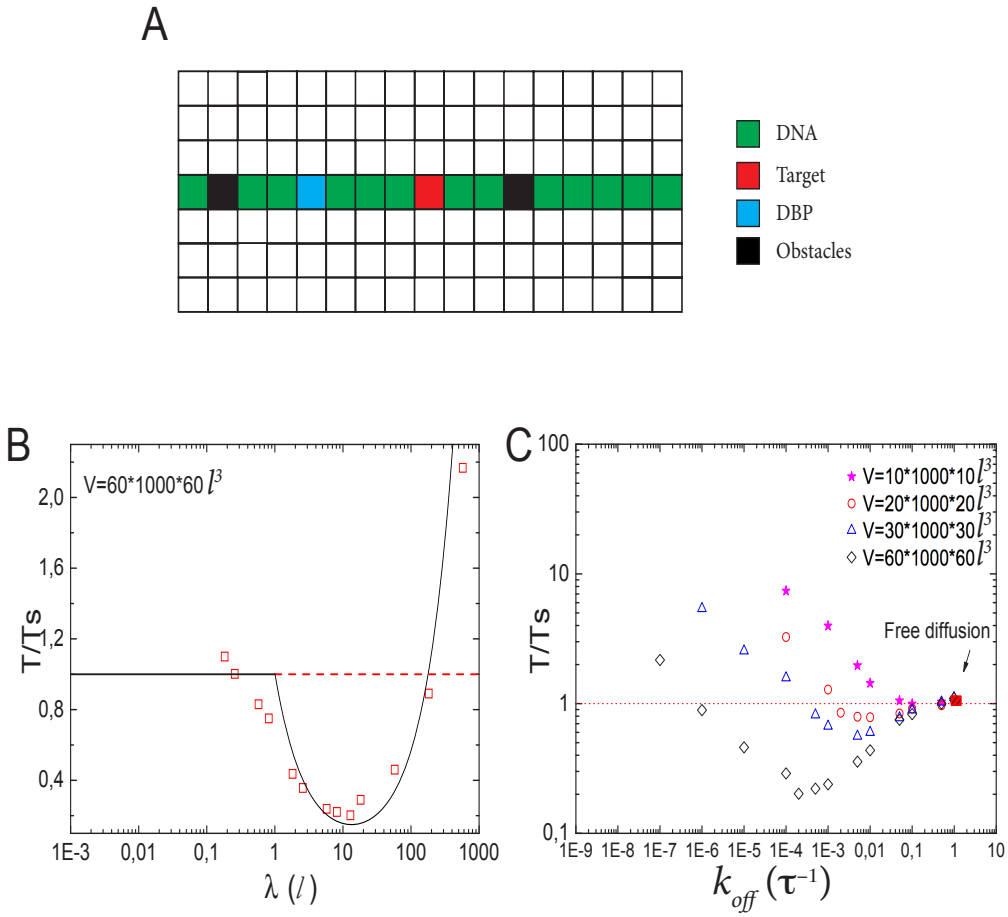

Figure 3.1 – Lattice model for facilitated diffusion: (A) Schematic view of the model. DBP (light blue) diffuses by hopping either to one of the six nearest neighbor lattice sites (3D) or to one of the two neighbors lattice sites if it is sliding on the DNA (1D), which is represented by the green line of lattice sites. O obstacles (black) are placed on the DNA. Their dynamics depend on the type of obstacle (see text), but in all cases they are not allowed to perform excursions to the bulk solution. Molecules have square geometry and diffusion is only allowed if the destination site is unoccupied. Note that the simulations use a three-dimensional lattice rather than the two-dimensional one depicted schematically here. (B) Search time (normalized by the search time T_s in the absence on non-specific binding of the DBP to the DNA) as a function of the sliding length λ. Data points are simulation results, and the line is the ratio between Eq. (3.5) and $T_s = V/4Da$. (C) T/T_s for different DNA concentrations, plotted as a function of the unbinding rate k_{pff}. Different minimal search times and optimal unbinding rates are observed for different DNA concentrations.

3.2 Lattice model for simulating facilitated diffusion

Facilitated diffusion is studied using Monte Carlo simulations of particles on a 3D lattice with periodic boundary conditions. The simulation box has a total volume $V = m_x \times m_y \times m_z l^3$, where l is the lattice spacing, chosen as the linear extension of the smallest particle type. A DNA molecule is implemented as a linear arrangement of lattice sites along the y axis with $x = z = (m_x/2)l = (m_z/2)l$. The system contains a target (e.g., a promoter or a regulatory binding site) and two types of particles: One DBP searching for the target and obstacles on the DNA, see Fig. 3.1A. All particles are taken to occupy exactly one lattice site. We take the target to be static, and its position is randomly chosen among the DNA lattice sites. The DBP is initially placed randomly in the bulk solution, whereas the obstacles are placed on DNA lattice sites, occupying a fraction of DNA equal to $O/\mathcal{L}$, where O is the number of obstacles and $\mathcal{L} = m_y$ the length of the DNA template. Here, obstacles are considered to be only on the DNA and they are not allowed to unbind and diffuse in the bulk solution.

In our implementation, the DBP can either diffuse in the bulk solution or slide on the DNA. If the DBP is in the bulk solution, at each simulation time step (of duration τ), it moves to each neighbor site with probability $1/6$. Thus, DBP diffuses in the bulk solution with diffusion constant $D_3 = l^2/(6\tau)$. If one of the neighboring sites is part of the DNA and it is free, the move is always accepted, and the DBP is placed onto the DNA. On the contrary, if the DNA site is occupied by an obstacle, the move is rejected and the DBP stays in the bulk solution (since obstacles are not allowed to unbind, moves within the bulk solution are always unhindered). Thus, no interactions other than steric repulsion are considered between the DBP and the obstacles.

If the DBP is bound to the DNA, it can either diffuse on the DNA or unbind from it with a rate k_{off}. If the DBP diffuses on the DNA, it performs a 1D random walk along the y axis with diffusion constant D_1. Unless stated otherwise, we choose $D_1 = D_3/10$, as 1D diffusion is typically slower than 3D diffusion (32; 71). Since we are interested in the effects that obstacles on the DNA have on the facilitated diffusion process, we implement three types of obstacles with different dynamics on the DNA: (i) static (immobile) obstacles, (ii) obstacles that diffuse on the DNA with a diffusion constant D_1, and (iii) obstacles that both diffuse and unbind from the DNA. Facilitated diffusion is expected to be affected differently by these three types of obstacles. 1D diffusion of both the DBP and the mobile obstacles is implemented as follows: If the destination site of a move is already occupied by another molecule, the move is rejected and the position is not updated. Position update of different molecules is performed in a random-sequential fashion: In every simulation step, we update the DBP position together with O randomly chosen obstacles, in such a way that on average all particles are updated once per simulation step.

1D diffusion takes places until the DBP unbinds from the DNA with probability P_{off}. Thus, P_{off} determines the rate at which the DBP unbinds from the DNA, $k_{\text{off}} = P_{\text{off}}\tau^{-1}$. We note that

experimentally, the unbinding rate is typically varied via the salt concentration (151; 152), because non-specific binding to DNA is typically of electrostatic nature and can be screened by large salt concentrations, resulting in large values of P_{off}.

In our simulations, the DBP finds the target either by 3D diffusion, if hopping from the bulk to one of the six neighbor lattice sites of destination happens to be the target, or by 1D diffusion if one of the two neighbor lattice sites is the target. Once the target has been found, we record the search time, and the DBP is randomly placed in the bulk solution.

3.3 Facilitated diffusion without obstacles

We start by considering the reference case of facilitated diffusion without obstacles on the DNA. This has been extensively studied in the past, both theoretically and computationally (28; 75; 78; 76; 79; 80; 81; 82; 150). In particular, it is well known that the efficiency of facilitated diffusion strongly depends on DNA concentration, D_1 and salt concentration (k_{off}) (33; 75; 155). The lattice model allows us to modulate these parameters and also provides a rather intuitive interpretation of the facilitated diffusion mechanism.

To begin with, we simulate facilitated diffusion in lattices of different sizes, thus effectively varying the DNA concentration. We keep the length of the simulation box (the direction parallel to the DNA) fixed at $m_y = 1000l$, but vary the box width in the other two dimensions ($m_x = m_z = L = 10, 20, 30$ and $60l$). Simulations are run until 2000 target finding events have taken place. In order to quantify the efficiency of facilitated diffusion in comparison to the free 3D diffusion case, we consider the ratio of the average search times T/T_s. Here, T is the average time for finding the target with facilitated diffusion, and T_s is the average time for finding the target via free 3D diffusion in the absence of DNA. Thus, if this ratio is less than one, facilitated diffusion is more efficient than the free diffusion case, whereas if the ratio exceeds one, it would be better for the system to find the target by simple 3D diffusion.

The essence of facilitated diffusion can be understood based on the following simple argument: The rate for target finding is diffusion-limited and thus given by $k = \gamma D a$, and the corresponding search time $T = (kc)^{-1}$, where D is the diffusion coefficient, a the size of the target, and $c = 1/V$ the concentration of the DBP (in our scenario a single one). γ is a numerical prefactor, which is $\gamma = 4\pi$ in the classical Smoluchowski result (36), and $\gamma = 4$ for our lattice model, where the target is a site on the DNA. Thus, the average time for target finding is obtained as $T = 1/k = V/(\gamma D a)$. The presence of the DNA has two different effects on the DBP dynamics, which can be both understood by introducing effective parameters: On one hand, transient binding to the DNA effectively slows down diffusion. Since diffusion along the DNA is typically slow, one can interpret binding to the DNA as pauses in 3D diffusion. These pauses can be incorporated into an effective diffusion coefficient

$$D_{\text{eff}} = D_3(1 - P_b), \tag{3.1}$$

where P_b is the probability for the DBP to be bound to the DNA. For the lattice model, this probability is given by

$$P_b = \frac{1}{1 + \frac{(L^2 - l^2)k_{\text{off}}}{4D_3}}. \tag{3.2}$$

On the other hand, sliding along the DNA effectively increases the size of the target. To reach the target, the DBP does not need to bind to the target site directly, but may also bind close to the target and slide there by 1D diffusion. The average distance, over which the DBP slides along the DNA while bound is given by

$$\lambda = \left(\frac{2D_1}{k_{\text{off}}}\right)^{1/2} \tag{3.3}$$

and is called the sliding length. Thus, the target size is effectively increased to $a_{\text{eff}} = \lambda$.

Taken together, the two effects lead to

$$\begin{aligned} k &= \gamma D_{\text{eff}} a_{\text{eff}} = \gamma D_3 (1 - P_b)\lambda \\ &= \gamma D_3 \frac{1}{1 + \frac{4D_3}{[(L^2 - l^2]k_{\text{off}}}} \times \sqrt{2D_1/k_{\text{off}}} \end{aligned} \tag{3.4}$$

and a search time of

$$T = (kc)^{-1} = \frac{V}{\gamma D_3 \lambda} + \frac{2L_c\lambda}{\gamma D_1}. \tag{3.5}$$

In the last expression, we have used $c = 1/V$ and defined the contour length of the DNA $L_c = V/(L^2 - l^2)$ to bring it into the form derived by Halford and Marko (76), with which the last expression agrees up to the numerical prefactor γ. It is important to notice that only the latter of the two effects facilitates diffusion to the target, while the former slows it down, so the name 'facilitated diffusion' is only appropriate in a limited range of the parameter space. Since measured 1D diffusion coefficients are typically smaller than 3D diffusion coefficients (32; 71), the slowing of diffusion is an important effect of non-specific binding to DNA and may well be the dominant effect for many proteins. Fig. 3.1B shows the ratio T/T_s as a function of λ for constant D_1. As λ increases, longer sliding events on the DNA take place when the DBP binds non-specifically to the DNA. Thus, the effective target size increases and facilitated diffusion becomes optimal at $\lambda \simeq 11l$. For larger λ, the DBP spends more and more time non-specifically bound to the DNA, thereby D_{eff} becomes smaller and facilitated diffusion becomes inefficient. Good agreement between our simulation data (red squares), and our analytical description (solid line) is observed. We note that the analytical argument given above is only valid for sliding lengths that exceed one lattice site. To correct for this effect, we have used an effective target size $a_{\text{eff}} = \max\{a, \lambda\}$ for the solid line.

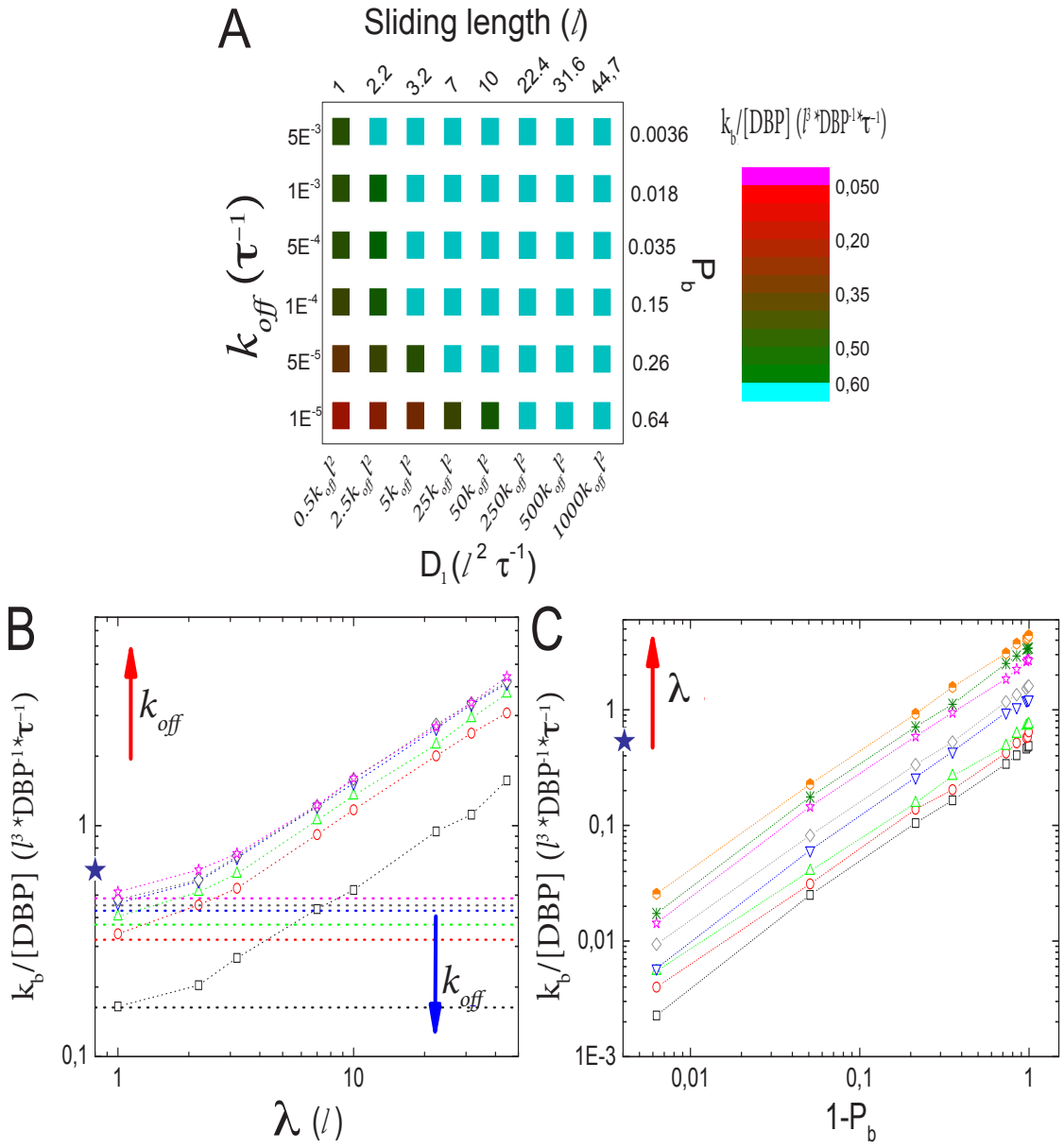

Figure 3.2 – Facilitated diffusion without obstacles on the DNA. (A) Target finding rate as a function of the two key parameters P_b and λ. Alternative axes show the microscopic parameters that are varied: k_{off} and D_1. k_{off} is varied with D_1/k_{off} constant. Light blue data represent the set of parameters that leads to finding rates faster than the free 3D diffusion process. (B) Promoter finding rate as a function of λ for different values of k_{off}. For sufficiently large λ, the finding rate is linear in λ and facilitated diffusion becomes more efficient than the free 3D diffusion process. Lines represent the limiting case where $D_1 = 0$, e.g. the DBP does not perform 1D diffusion on the DNA. (C) Finding rate as a function of the probability for the DBP to be unbound from the DNA $(1 - P_b)$ for different values of λ. For low values of $(1 - P_b)$, the finding rate increases linearly, in agreement with $k_b/[L] \propto D_{\text{eff}} = D_3(1 - P_b)$.

An important feature of this result is that the search time is non-monotonic as a function of the DNA binding strength. This feature has been a hallmark of facilitated diffusion, which has been at the center of many theoretical studies (69; 70; 76; 80; 82), and which was confirmed experimentally by varying salt concentration (151; 152). Based on our simple argument above, it can be interpreted as arising from the opposite effects of increasing the strength of non-specific binding on the effective diffusion coefficient and the effective target size.

Fig. 3.1C shows T/T_s as a function of k_{off} for four different DNA concentrations. In all the cases, a non-monotonic behavior is observed as the value of k_{off} is modulated. For low values of k_{off}, ratios larger than one are obtained, reflecting mostly the effective reduction in the 3D diffusion constant. Thus, in this parameter range, 3D diffusion in the absence of DNA is more effective that facilitated diffusion. For intermediate values of k_{off}, a minimum in the ratio T/T_s is observed. In this case, sliding on the DNA indeed facilitates binding to promoter, due to the effectively larger target size, i.e. $\lambda + 1$. For high values of k_{off}, the ratio T/T_s gets close to one, as DNA binding is weak and only has a small effect.

For the case $V = 60 \times 1000 \times 60l^3$, the optimal value of k_{off} is found to be $0.0002\tau^{-1}$. Then, the DBP diffuses on average over a sliding length of $\lambda \simeq 11l$. With the obtained optimal value of k_{off} we can obtain the probability that the DBP is bound to the DNA $P_b = 0.48$, see Eq. (3.1). Thus, on average, the DBP spends approximately the same time bound to the DNA as diffusing freely in the bulk solution. This is in agreement with Marko's model, which predicts that the fraction of time the DBP spends on the DNA is the same fraction of time it spends in the bulk solution (76). Notably, the latter prediction is not general and for cases with high DNA concentration it is not longer valid. For example, when taking the case $V = 10 \times 1000 \times 10l^3$, we find that the optimal unbinding rate is $k_{\text{off}} = 0.1\tau^{-1}$. When plugging the latter value of k_{off} into Eq. (3.2), we obtain that $P_b = 0.063$. Here, the very frequent DBP-DNA non-specific binding events drastically reduce D_3 if the DBP spends relatively large times on the DNA. Thus, the optimal DBP dynamics is the one in which, the DBP spends most of the time in the bulk solution with very short sliding events on the DNA, see the pink data in Fig. 3.1B.

Though similar qualitative behavior in facilitated diffusion is observed for all DNA concentrations, large quantitative differences are obtained. As DNA concentration increases, the minimum in the ratio T/T_s gets close to one, and the optimal k_{off} value shifts to the right. This shows that for large DNA concentrations, it is more efficient for the system to tune the k_{off} parameter to large values, and thus, decrease P_b.

Eq. (3.5) suggest that there are two relevant parameters of facilitated diffusion, the binding probability P_b and the sliding length λ. Each of these parameters determines the strength of one of the two effects of non-specific DNA binding. They are related to the microscopic parameters k_{off} and D_1, but using these two parameters does not separate the two effects.

Experimentally, a modulation of the parameters via salt concentration modulates k_{off} and thereby acts via both effects.

As a check of the parameter dependencies indicated by Eq. (3.5), we systematically vary P_b and λ. This is done microscopically by modulating k_{off} with the ratio D_1/k_{off} fixed and D_1 for constant k_{off}, respectively. The results are shown in Fig. 3.2. The color map in Fig. 3.2A shows the rate for target binding as a function of both parameters. Light blue color indicates parameter combinations for which 'facilitated diffusion' is indeed optimal, i.e. the combination of 3D and 1D diffusion is more efficient than free 3D diffusion ($T/T_s < 1$). It is rather inefficient if the DBP remains bound to the DNA for long time without scanning the DNA. The dependencies on the individual parameters are shown in Fig. 3.2B and 3.2C, respectively. Fig. 3.2B shows the promoter finding rate as a function of the sliding length for constant values of k_{off} and, thus, P_b. The horizontal lines represent the limiting case where the DBP binds to the DNA but does not diffuse along it ($D_1 = 0$). The star indicates the finding rate in the absence of DNA. As expected from Eq. (3.5), the rate increases linearly with increasing sliding length (except for sliding lengths $\lambda \lesssim l$, where the effective target sizes becomes $\simeq$ l). The linear increase is seen for all values of k_{off}. The plot also shows that sufficiently large sliding lengths are needed in order for the binding rate to exceed the free binding rate in the absence of DNA. The minimal required sliding length for facilitation increases with increasing P_b or decreasing k_{off}.

Fig. 3.2C shows the finding rate as a function of $(1 - P_b)$, the probability that the DBP is free in solution, for constant sliding length. For low values of $(1 - P_b)$, the DBP frequently binds non-specifically to the DNA. These events effectively reduce the 3D diffusion constant to $D_{\text{eff}} = D_3(1 - P_b)$. In agreement with this expectation, the finding rate increases linearly with the probability to be unbound, $1 - P_b$.

3.4 Effect of obstacles on the DNA

Next, we consider the impact that obstacles on the DNA have on facilitated diffusion. To that end, we implement three types of obstacles with different dynamics: Obstacles of the first type diffuse on the DNA, but do not unbind. We set the crowders to diffuse on the DNA with the same 1D diffusion constant D_1 as the DBP. The obstacles can transiently occupy the target position, making it inaccessible for the DBP, independent of whether the DBP arrives via 3D diffusion or 1D sliding. The second type of obstacles that we consider are static (D_1=0), with obstacles regularly spaced on the DNA. For this case, we make sure that the target is placed on a free DNA lattice site between the obstacles. Obstacles of type three also diffuse on the DNA, again with the same diffusion constant as the DBP, but, in addition, these obstacles can dissociate from the DNA with a rate $k_{\text{off}}^{\text{Obst}}$. In these simulations, we do not track diffusion of unbound obstacles in the bulk. Rather, when an obstacle dissociates from the DNA, we

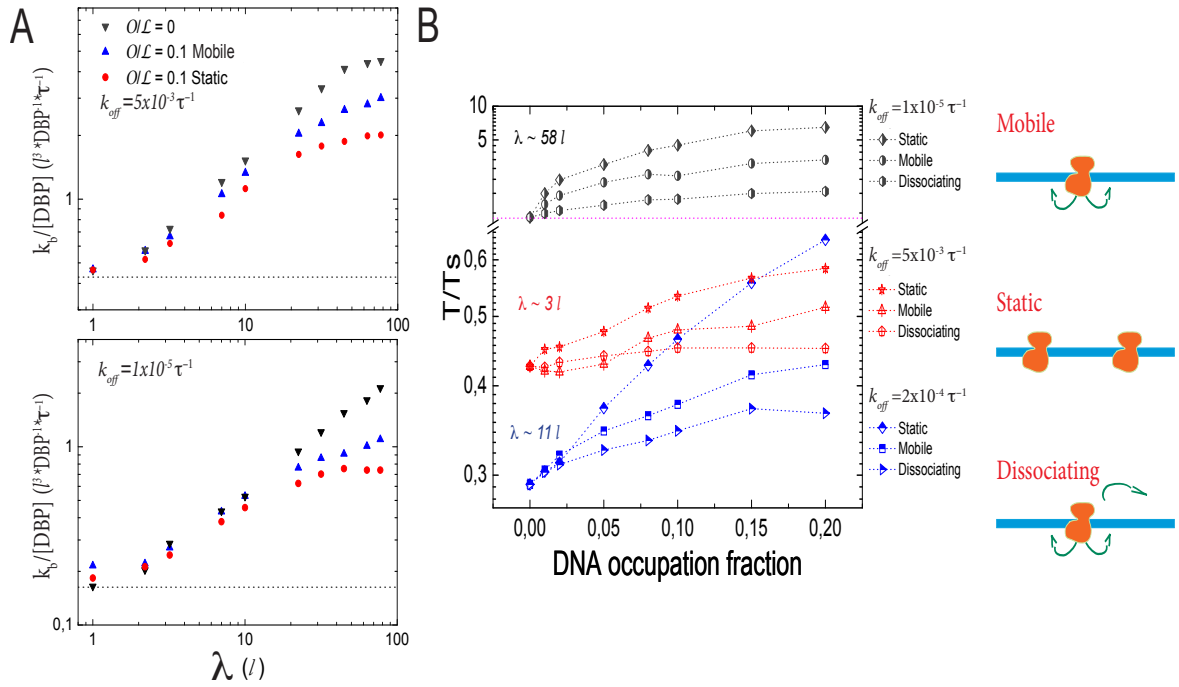

Figure 3.3 – Effects of obstacles on the DNA on facilitated diffusion. (A) Promoter finding rate as a function of λ for different values of k_{off} in the presence of obstacles on the DNA. In general, presence of obstacles on the DNA hinders promoter finding. Depending on the type of obstacles, the binding rate is affected in a stronger or more moderate way. Lines represent the limiting case $D_1 = 0$, where the DBP does not diffuse on the DNA. (B) Dependance on obstacle coverage: Search time ratio T/T_s, as a function of the occupation fraction for three types of obstacles. Mobile obstacles perform 1D diffusion on the DNA, static obstacles are regularly distributed on the DNA and dissociating obstacles diffuse on the DNA and dissociate from it with an unbinding rate $k_{\text{off}}^{\text{Obst}} = 0.0002\tau^{-1}$. Static obstacles have the strongest effects on facilitated diffusion, whereas obstacles that dissociate from the DNA have the smallest.

immediately place a new obstacle on a random position of the DNA in order to keep the DNA occupation fraction constant, see representation of the obstacles in Fig. 3.3. The obstacle that dissociates is removed from the simulation.

Based on the two effects of (non-specific) binding to DNA discussed above, one can argue that obstacles on the DNA affect target finding only through the effective target size. By contrast, obstacles or crowders in the solution, which we do not consider here, would mostly affect it via the effective diffusion, although one can also expect an increase in the target size (due to crowding-enhanced binding to the DNA, which increases the sliding length (28; 75)). Obstacles on the DNA that act as barriers for sliding effectively reduce the sliding length and thus the target size. Thus one can estimate an effective target size as the smaller one of the two length scales of this system, the sliding length on one hand, and the average distance between obstacles on the other.

However, Fig. 3.3A shows that the dynamics of the roadblock also matter for the results. Here, we run simulations for two different values of k_{off}, $k_{\text{off}} = 0.005\tau^{-1}$ and $0.00001\tau^{-1}$, at 10% DNA occupation. Shown data are for obstacle types one and two. The plot shows the target finding rate as a function of λ when modulating D_1. Lines represent the limiting case of $D_1 = 0$. As D_1 increases, the finding rate becomes larger. For $\lambda \simeq 3l$ the promoter finding rate starts to differ between the different types of obstacles. On one hand, at $O/\mathcal{L} = 0.1$ there is an obstacle every 10 lattice sites if they are static. Thus, as D_1 increases, the search scenario becomes effectively one in which the DBP finds a target of size $a_{\text{eff}} = 10l$, with an effective three dimensional diffusion constant D_{eff}. On the other hand, if obstacles diffuse on the DNA, a_{eff} fluctuates and can transiently be reduced or increased. Thus, saturation takes place at larger values of λ. We note that this average size depends on the obstacles' one-dimensional diffusion constant.

In Fig. 3.3B we plot the ratio T/T_s as a function of the fractional occupation of the DNA by obstacles, for three different values of k_{off}, $k_{\text{off}} = 0.005, 0.0002$ and $0.00001\tau^{-1}$. These values correspond to sliding lengths of $\lambda \simeq 3l, 11l$ and $58l$, respectively. We note that for all cases, the presence of obstacles has a negative effect on facilitated diffusion, even for intermediate values of k_{off} at low occupation fractions.

In general, static obstacles have the largest impact on facilitated diffusion, because they effectively trap the DBP once it has been bound non-specifically to the DNA. In this case the only way that the DBP can overcome an obstacle is to unbind from the DNA. Thus, the negative effect of static obstacles on the ratio T/T_s becomes larger as the unbinding rate k_{off} decreases. Type-1 obstacles have the second largest effect on facilitated diffusion. Here, even though obstacles trap the DBP in regions where the target may not be present, this trapping changes as the obstacles diffuse on the DNA, making the target accessible to the DBP, even if it remains bound to the DNA for long time. However, 1D diffusion is effectively slowed down by obstacles on the DNA. The effect of the obstacles decreases as the DBP unbinding rate

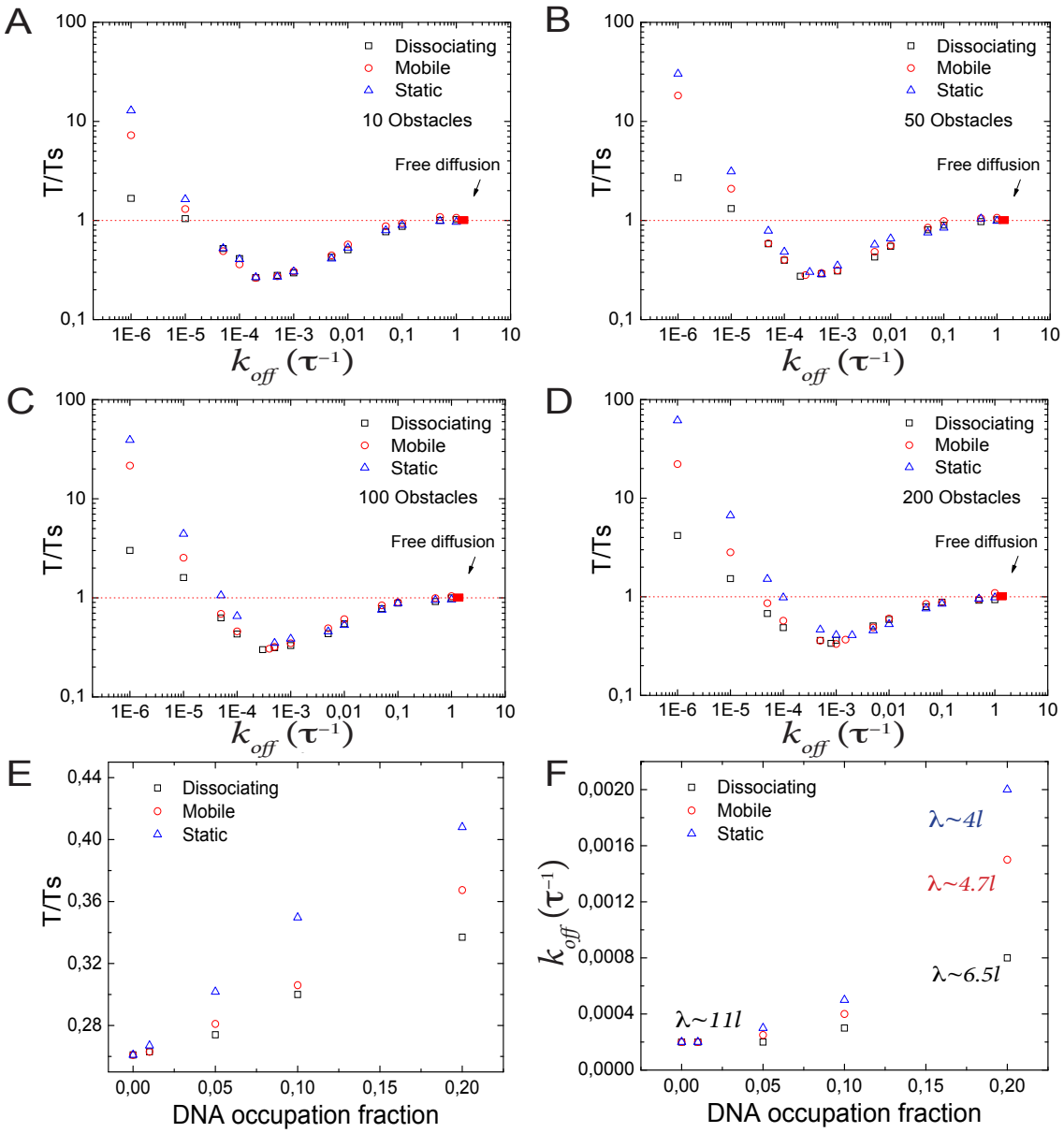

Figure 3.4 – Facilitated diffusion in the presence of obstacles on the DNA. (A-D) Search time ratio T/T_s as a function of k_{off} for the three types of obstacles. As the DNA occupation fraction increases, the minimum in T/T_s change, as well as the value of k_{off} where it occurs. (E) Minimal search time as a function of the DNA occupation fraction. As the number of obstacles increases, facilitated diffusion becomes less efficient, with the static obstacles affecting facilitated diffusion the most. (F) Corresponding optimal value of k_{off} as a function of the DNA occupation fraction. In order to be as efficient as possible, the DBP must spend more time in the bulk solution, to avoid being trapped between obstacles.

increases. Type-3 obstacles have the weakest effect on facilitated diffusion. In this case, the DBP can access the target in the same way as for type one. Since the obstacles can unbind from the DNA, the trapping effect is weaker and the DBP can scan larger sections of DNA. Thus, the impact of this type of obstacles on facilitated diffusion is more modest.

Interestingly, for different types of obstacles and DBP unbinding rates, the ratios T/T_s can get close and even cross (Fig. 3.3B). To take a better look into this issue, we reconstruct the facilitated diffusion plot for different levels of fractional DNA occupation, and obtain the k_{off} at which facilitated diffusion is optimal, see Fig. 3.4A-D. As can be observed, the minimum shifts towards higher values of k_{off} and higher values of T/T_s. Next, we plot in Fig. 3.4E the optimal values of the ratio T/T_s as a function of DNA occupation. Static obstacles have the strongest effects on facilitated diffusion, and the dissociating obstacles the weakest. In Fig. 3.4F, we plot the corresponding values of k_{off} at which facilitated diffusion is the most effective. Here, as the DNA occupation increases, the DBP needs to bind less tightly and scan smaller sections of DNA at every 1D sliding excursion. This suggest that in order to be as efficient as possible, the DBP has to spend more time in the bulk solution than on the DNA. Specifically, we showed above that in the absence of obstacles on the DNA, the DBP spends on average $\simeq 50\%$ of the total search time bound to the DNA. In the presence of static obstacles occupying 20% of the DNA, the average time the DBP spends bound to the DNA under optimal conditions is decreased to only 9%.

3.5 Concluding remarks

In this paper, we used a simple computational approach to study the promoter finding process via facilitated diffusion. We used lattice simulations to revisit how the promoter finding process is affected when modulating different parameters in our simulations. Non-specific binding to the DNA effectively reduces diffusion in the bulk and thereby the diffusion-limited binding to the target. At the same time, sliding along the DNA effectively increases the size of the target. The competition of these two effects leads to a characteristic maximum in the binding rate. This maximum is very sensitive to the concentration of DNA, D_1, k_{off} and λ. The lattice model provides a rather intuitive picture of these situations as well as an efficient computational implementation of facilitated diffusion.

In addition, we have studied the effects that different types of obstacles have on the facilitated diffusion process. In general, since the DBP is (at least transiently) trapped between two obstacles, its sliding length is reduced, decreasing effectively the target size. Beyond that, our results suggest that obstacles with different dynamics on the DNA have different effects on facilitated diffusion, with static obstacles affecting the finding rate the most. In order to be as efficient as possible, the DBP has to modulate its interaction with the DNA, and spend more time in the bulk solution. The methods we used here can be applied more

generally and there are a number of possible extensions to this work, such as addressing the effects of crowders on the bulk solution on the facilitated diffusion process. Other aspects of facilitated diffusion will require more detailed molecular approaches, for example the effect of DNA flexibility and supercoiling, which are difficult to implement in the lattice model. The latter has been shown to have an effect on target search kinetics with more rapid target finding for coiled DNA conformation than for stretched DNA conformations (77; 93; 156). Studying the interplay of DNA conformation and obstacles will be a task for future work.

Chapter 4

Mechanism of facilitated diffusion during DNA search in crowded environments

Abstract

The key feature explaining the rapid recognition by a protein of its DNA target site lies in the combination of one- and three-dimensional (1D and 3D) diffusion, which allows efficient scanning of the many alternative sites. This facilitated diffusion mechanism is expected to be affected by cellular conditions, particularly crowding, given that up to 40% of the total cellular volume may by occupied by macromolecules. Using coarse-grained molecular dynamics and Monte-Carlo simulations, we show that the crowding particles can enhance facilitated diffusion and accelerate search kinetics. This effect originates from a trade-off between 3D and 1D diffusion. The 3D diffusion coefficient is lower under crowded conditions, but it has little influence because the excluded volume effect of molecular crowding restricts its use. Largely prevented from using 3D diffusion, the searching protein dramatically increases its usage of the hopping search mode which results in higher linear diffusion coefficient. The coefficient of linear diffusion also increases under crowded conditions as a result of increased collisions between the crowding particles and the searching protein. Overall, less 3D diffusion coupled with an increase in the usage of hopping and speed of 1D diffusion results in faster search kinetics under crowded conditions. Our study shows that search kinetics and mechanism are modulated not only by crowding occupancy, but also by the properties of the crowding particles and the salt concentration.

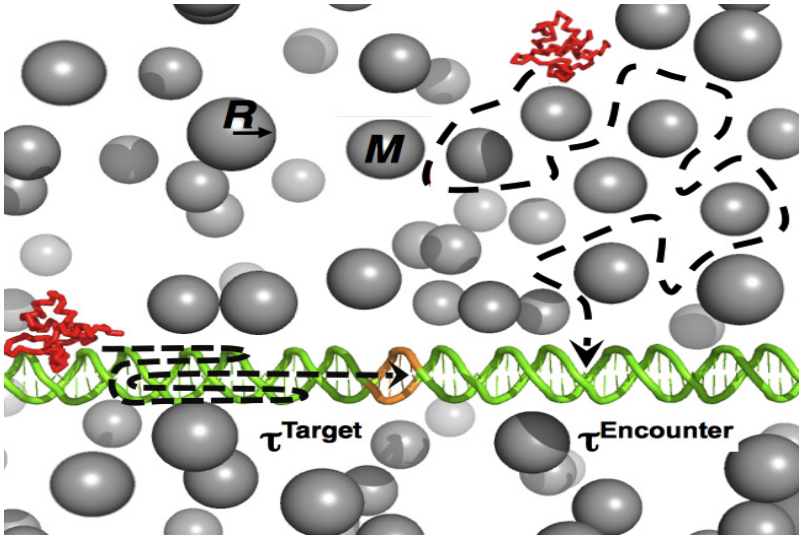

Figure 4.1 – Schematic of DNA search by a protein in the presence of crowding particles. The DNA-binding protein and the spherical crowding particles are shown in red and grey, respectively. The double-stranded DNA is shown in green and the target site in orange. The crowding condition is characterized by the volume fraction of the crowding particle, φ, which is affected by the number of particles, N, and their radius, R. The mass of the particle, m, may also affect the overall crowding effect. Crowding may affect the characteristic time to localize the target site, τ^{Target} (dashed blue line) and the time to reach the DNA when starting from a position in the bulk, $\tau^{\mathrm{Encounter}}$ (dashed black line). The crowding condition shown in the figure corresponds to a fractional volume of φ=50%.

4.1 Introduction

The phenomenon of specific molecular recognition between two biomolecules lies at the heart of many biological processes. In the case of interactions between a DNA-binding protein (DBP) and its specific target DNA sequence, recognition requires that the DBP locates its target site among many alternative sites that have a sequence similar to that of the specific target site. Clearly, a failure to rapidly find the target site may result in cellular malfunction, because these binding events are often part of a cascade of various essential events in which timing is critical.

The experimentally observed speed of the DNA target search conducted by DBPs can only be resolved by considering a search mechanism involving facilitated diffusion in which the DBPs find their target sites on the DNA through a combination of one- and three- dimensional (1D and 3D) diffusion (69; 82; 152; 155; 157). 1D diffusion includes relatively short bi-directional random walks during which the DBP performs a coupled rotation-translation motion as it moves along the linear contour of the DNA (‘sliding’) combined with short-range jumps between neighboring DNA segments (‘hopping’). 3D diffusion events involve dissociation from the DNA to the bulk followed by reassociation, thus allowing the DBP to visit DNA regions that are sequentially distant. The physics and biochemistry of the facilitated diffusion mechanism have been investigated from both the theoretical and experimental perspectives,

including through complex kinetic models, simulation tools (both at the coarse-grained and atomistic levels) (155; 158; 159; 160; 161; 162), and in vitro biochemical measurements at the bulk (163; 164; 165; 166; 167) and single molecule levels (32; 168). The in vivo cellular environment, which is influenced by DNA packing, the positions of nucleosomes, high concentrations of metabolites and macromolecules, and the involvement of other proteins in many regulatory processes, is complex and may directly and indirectly affect search kinetics and mechanisms. This cellular complexity may dynamically affect the biophysics of facilitated diffusion in various ways. Although some aspects of DNA search in vivo have been investigated experimentally (4) and some from the theoretical perspective (169; 170), much remains to be understood, particularly in the context of the crowded cellular environment (171).

The search dynamics, which is governed by nonspecific protein-DNA interactions, is dominated by electrostatic forces and is therefore strongly dependent on the salt concentration, as this influences partitioning between the 1D and 3D search modes of the facilitated diffusion process. At a low ionic strength, the electrostatic interactions between the DBP and DNA are stronger, and the protein mostly diffuses along the DNA major groove using a relatively short bi-directional 1D random walk. As the salt concentration increases, the protein may move away from the DNA and diffuse into the 3D bulk, or remain in the vicinity of the DNA and undertake a linear search via hopping, which is accompanied by a higher diffusion coefficient and enhanced DNA scanning (155). However, such a search mechanism takes place in a crowded cellular environment in which additional macromolecules may be adsorbed onto the DNA and impose a physical constraint on the search process (4; 6; 7; 10; 22; 124), as shown in Fig. 4.1. Given that the macromolecules occupy 10-40% of the total cellular volume (which corresponds to a concentration of $\sim$100-300 mg/mL (6; 123), they are expected to have a non-negligible effect on search kinetics. The macromolecular density of the bacterium *Escherichia coli*, for example, slows down 3D diffusion by an order of magnitude, and about 30% of its DNA is associated with proteins, which are mostly bound with no sequence specificity. Macromolecular crowding can, in principle, affect the kinetics of DNA search in various ways. For example, crowding involving large molecules can exclude some of the 3D volume from the purview of the smaller searching protein in what is known as the 'exclusion effect' (also called depletion forces). Macromolecular crowding can also act as obstacles and block the 1D dynamics of the searching protein along the DNA and thereby truncating its sliding track into short fragments (33). As many DBPs utilize a mixture of the 1D and 3D search mechanisms to facilitate localization of their DNA target (69), it is interesting to explore the effect of molecular crowding on the efficiency of the search and in particular how crowding interacts and potentially interferes with both the 1D and 3D components of DNA search. The effect of molecular crowding on the kinetics of protein-DNA interactions has shown that facilitated diffusion is affected by the cellular conditions (18; 172). In particular, a recent

study suggested that a crowded cellular environment could enhance the mechanism of DNA search. Apparently, by creating low-viscosity micro-environments around the enzyme and DNA, crowding increases the likelihood that the enzyme will successfully translocate between its respective target sites without dissociating into the bulk solution and also increases the average translocation distance. Furthermore, the protein can traverse a larger linear distance on the DNA chain in the presence of crowding (18). While the effect of crowding on protein-DNA recognition cannot be generalized because of the scarcity of data, it should be mentioned that crowding seems to have diverse effects on protein-protein recognition (173). Theoretical studies have suggested different arguments regarding the role of crowding in DNA search kinetics and efficiency (28; 88; 137). The lack of consistency in these studies reflects that the underlying physical principle governing the effect of crowding on protein-DNA interaction is unclear. Thus, despite considerable theoretical (82; 150; 169; 174; 175; 176) and experimental efforts (177), researchers still lack a detailed description and understanding of the possible effects that macromolecular crowding of 1D and 3D movements exerts on the overall efficiency with which a protein searches DNA (88).

This study seeks to understand the consequences of cellular crowding on DBP search mechanisms and kinetics. Coarse-grained molecular dynamics (CGMD) simulations of non-specific protein-DNA interactions have recently provided new mechanistic insights into the DNA search performed by various proteins (155; 158; 178; 179; 180; 181; 182; 183) and have captured key aspects of the search features observed experimentally, such as the speed-stability paradox (184; 185). Here, we utilized CGMD to elucidate the effect of crowding on DNA search by DBPs with the aim of furthering understanding of the complexity of target site recognition in the cell.

4.2 Results and Discussion

4.2.1 The mechanism of facilitation in DNA search: The effect of molecular crowding on 1D and 3D diffusion

CGMD simulations can provide insights into the molecular mechanism underlying enhanced DNA search by a DBP. In particular, the simulations may explain the complex effect of the degree of crowding, salt concentration, and enhanced search kinetics. The effect of crowding on the search dynamics can be elucidated from the search modes (i.e., sliding, hopping, and 3D diffusion) adopted under each condition. Previous studies showed that the search efficiency of a DBP is largely dependent on salt concentration (155). In this study, we examined how crowding can further affect the search mechanism adopted by the DBP. To address this, we simulated a DBP searching a ds-DNA molecule at salt concentrations ranging from 0.01-0.09 M in the presence of crowding macromolecules that had only an entropic effect (i.e., their interactions with the protein or with the DNA were modeled solely by excluded

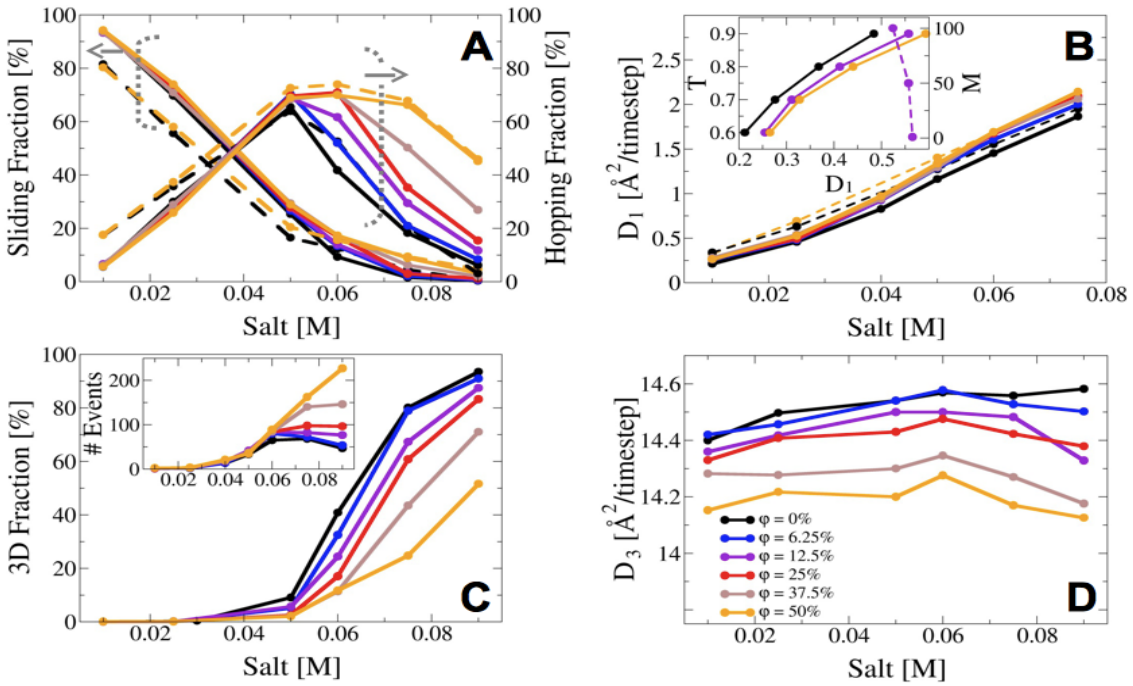

Figure 4.2 – The effect of molecular crowding on the mechanism of DNA search by proteins. A) The effect of crowding on the usage of the different types of 1D diffusion (i.e., sliding and hopping, as identified by the dotted brackets and associated arrows). B) The linear diffusion coefficients (D_1) are shown for different salt concentrations and fractional volumes (φ) of crowding particles. The dotted lines in A) and B) correspond to protein dynamics along flexible DNA with φ=0 (dotted black) and φ=50% (dotted orange), whereas the sold lines in all four sub-figures represent protein dynamics along fixed DNA. The inset in (B) shows the effect of the temperature (solid lines) and mass (dotted line) of the crowder particles on the values of D_1 for different φ conditions. C) The propensity of the protein to search using 3D diffusion under different salt concentrations and for different fractional volumes and (inset) the number of dissociations from 1D to 3D. D) The D_3 diffusion coefficient in the bulk for different salt concentrations and factional volumes.

volume). Crowding simulations were performed for various fractional volumes (φ), ranging between 0-50%, where the value of φ is a function of the number of crowding particles, N and their size, R. As control simulations, we studied and compared the search mechanism by the protein in the absence of crowding macromolecules (i.e., φ =0%). Figure 4.2A illustrates that, irrespective of the crowding conditions, the usage of the sliding search mode decreases with increasing salt concentration, while (at low salt concentrations) that of hopping increases. Interestingly, the utilization of both search modes is unaffected by crowding at relatively low salt concentrations between 0.01-0.05 M NaCl even for a high volume fraction of 50%. At higher salt concentrations ([NaCl]>0.05 M), the proportion of sliding adopted by the searching protein is slightly affected by crowding, with more sliding observed at φ=50% compared with φ=0%. A more profound effect of crowding is observed at higher salt concentrations with respect to the hopping search mode, with a considerably higher proportion of hopping events occurring at φ=50% compared with φ=0%. The increase in sliding dynamics and especially in hopping dynamics associated with greater crowding occurs at the expense of 3D diffusion, which decreases as crowding increases. For example, upon changing φ from 0% to 50%, the propensity for 3D diffusion reduces from 93% to 52% (Fig. 4.2C). The sharp reduction in productive DBP-DNA dissociation can be attributed to the volume exclusion effect of the crowder molecules, which restrict the ability of the DBP to escape from the DNA and hinder the performance of a 3D excursion in the bulk. When such a 3D event occurs, it is expected to be much shorter for higher φ values because the high viscosity of the crowders will prevent the protein from traversing in solution. Consequently, greater crowding results in greater protein-DNA affinity (28).

An increase in the number of hopping events under conditions of greater crowding occurs only at moderate and high salt concentrations (Fig. 4.2A), yet faster search kinetics as a consequence of crowding is also observed at a lower salt concentration (Fig. 4.3A). The faster search enabled by crowding at low salt concentrations stems from a different mechanism, which is suggested because in this range of salt concentrations the relative usage of the three search modes (i.e., sliding, hopping and 3D) is very similar. We therefore further analyzed the physical characteristics of the linear diffusion component, which is the most dominant search mode at this salt concentration. Figure 4.2B shows the 1D diffusion coefficient (D_1) at various fractional volumes under salt concentrations less than 0.08 M (above such salt concentrations, the protein performs mostly 3D diffusion in the bulk). As can be seen, introducing molecular crowding results in an increase in the D_1 diffusion coefficient at all salt concentrations. Consistently with the pronounced increase in the proportion of hopping with increased crowding (Fig. 4.2A), this increase becomes greater at higher salt concentrations. Introducing flexibility to the DNA molecule (dotted lines in Figures 4.2A and 4.2B) reduced the absolute sliding fraction compared with the values obtained using a rigid DNA molecule and consequently affected the fraction of hopping too. However, DNA flexibility did not alter the effect of crowding or markedly change the shapes of the curves.

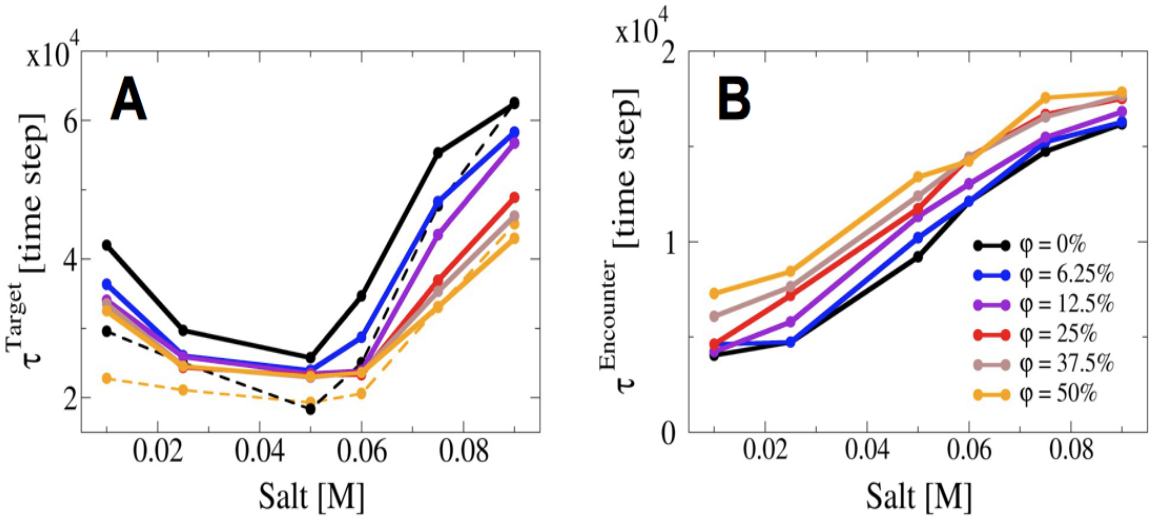

Figure 4.3 – The effect of molecular crowding (modeled as φ, the fractional volume) on the kinetics of DNA search. A) The mean time to approach the target site when starting from a selected position on the DNA is indicated by τ^{Target} (see Fig. 4.1) at different salt concentrations. The dotted lines correspond to protein dynamics along flexible DNA, with φ=0 (dotted black) and φ=50% (dotted orange), whereas the solid lines represent dynamics on fixed DNA. B) The mean time for the protein to encounter any nonspecific DNA site, $\tau^{\text{Encounter}}$, is measured by placing the protein far from the DNA at different salt concentrations.

To understand the origin for the increase in D_1 at low salt concentrations in the presence of crowding particles, we studied the effect of increasing the temperature or the mass of the crowding molecules on the linear diffusion coefficient. The inset of Figure 4.2B shows that D_1 increases with temperature irrespective of φ, but at higher values of φ, the increase in D_1 is larger. Nonetheless, we point out that changing the temperature does not affect the fraction of hopping. Increasing the mass of the crowders by a factor of 50 or 100 results in lower D_1 values. The response of the D_1 of the DBP to the simulation temperature and to the mass of the crowders suggests that the crowders affect D_1 regardless of a change in the population of the hopping searching mode. Accordingly, at low salt concentrations, the dependence of D_1 on temperature and mass suggests that the crowders constantly collide with the searching protein and consequently D_1 increases. Increasing the temperature or reducing the mass of the crowders increases the frequency of these collisions and further increases D_1. The higher D_1 in the presence of molecular crowding is reminiscent of the higher D_1 reported recently in the context of collisions between a DBP and flexible DNA (159).

We then examined the effect of molecular crowding on 3D diffusion in the bulk (D_3). Figure 4.2D shows that D_3 is independent of salt concentration for all values of φ. D_3 decreases as the fractional volume of molecular crowding increases. This suggests that, for large values of φ, the DBP is more confined in 3D and that its 3D diffusion is slower in this viscous environment.

Crowding causes D_3 to decrease by up to 2% (Fig. 4.2D) while it causes D_1 to increase by up to 15% (Fig. 4.2B).

4.2.2 Effects of crowding fractional volume on the kinetics of finding the DNA target site

Molecular crowding may affect the kinetics of DNA search by affecting the nature of 1D and 3D diffusion. The size of the effect is expected to depend on the fractional volume of the crowders (i.e., on φ) and on their properties, such as their dimension, mass, or affinity to DNA molecules or to proteins. The crowders are expected to affect both 1D and 3D diffusion. We therefore designed simulations to quantify independently these two effects. To estimate the effect of crowding on 1D diffusion, we measured τ^{Target}. The effect of crowding on 3D diffusion was estimated by $\tau^{\mathrm{Encounter}}$. The values of τ^{Target} and $\tau^{\mathrm{Encounter}}$ as a function of salt concentration and for different crowding conditions are shown in Figs. 4.3A and 4.3B, respectively. The time needed for a DBP to find its target site regardless of crowding volume fraction strongly depends on the salt concentration, with the fastest search (i.e., smallest τ^{Target}) achieved at a moderate salt concentration ($\sim$0.05 M, Fig. 4.3A). This salt concentration corresponds to the optimal condition for achieving the most efficient search in terms of the balance between the 1D and 3D search modes. At lower salt concentrations, searching is dominated by 1D diffusion along the DNA (because the relatively weak electrostatic screening that occurs as low salt concentrations promotes tighter nonspecific protein-DNA interactions), whereas at higher salt concentrations, 3D diffusion is much more common. Accordingly, at higher salt concentrations, a high search speed is achieved by a combination of the 1D and 3D search modes. Figure 4.3A shows that searches proceed faster in more crowded environments compared with less crowded ones under all salt conditions, although both the absolute value of τ^{Target} and the size of the difference between the τ^{Target} values for different fractional crowder volumes depend on the salt concentration. For example, while crowding at φ=6.25-50%, enhances the search rate by about 30% at low salt concentrations, at high salt concentrations the size of the effect depends on the exact volume fraction. For φ=6.25%, a mildly enhanced kinetic effect of about 8% is measured, whereas for φ=50% the effect is of 45%. This suggests that crowding may exert different effects on 1D dynamics compared with 3D dynamics and that these opposite effects do not necessarily cancel out (28). Nonetheless, we note that the dependency of τ^{Target} on salt becomes weaker with increasing φ, as reflected by the smaller gap between the search kinetics at salt condition extremes and salt condition for optimal search efficiency. This is another manifestation of a more robust search due to crowding (28).

We suggest that at moderate and high salt concentrations, the faster search kinetics observed with a higher volume fraction of crowding (Fig. 4.3A) is due to 1D diffusion, specifically hopping, increasing to compensate for 3D dissociation failures. Figure 4.3B shows that,

indeed, at high salt concentrations combined with close crowding, the low fraction of 3D is characterized by many short events while, in the absence of crowding, 3D diffusion comprises much fewer dissociation events that are very long. To highlight the influence that crowding exerts on DBP search kinetics and mechanism, we present (Figure A.1) six trajectories of protein dynamics at two crowding volume fractions: φ=0% (right panels) and φ=50% (left panels) and under three representative salt concentrations. These six trajectories illustrate the faster kinetics for binding the target site (lower values of τ^{Target}, indicated by the horizontal red arrow) for φ=50% compared with φ=0% and the stronger influence of crowding at higher salt concentrations. The trajectories also illustrate the lower percentage of 3D diffusion at higher crowder volume fractions. At both salt concentrations 0.06 M and 0.09 M, dissociation from the DNA molecule is repressed by the crowding.

To investigate potential crosstalk between molecular crowding and the properties of the DNA molecule, we also studied the effect of crowding on search kinetics using a model that captures DNA flexibility. We found that, in the absence of crowding, τ^{Target} is shorter and the search speed increases by about 30% when the DNA molecule is flexible rather than rigid, although the overall nature of the search remains unchanged (Fig. 4.3A, dotted lines). The small effect of DNA flexibility on partitioning between sliding/hopping/3D diffusion seen in our study is in accordance with some earlier reports (158; 186). The effect of crowding on the time required to locate the target site is similar for rigid and flexible DNA. Crowding also strongly affects $\tau^{\mathrm{Encounter}}$. In the absence of crowders, $\tau^{\mathrm{Encounter}}$ increases with increasing salt concentration (Fig. 4.3B). The slowing of the time required to achieve DBP-DNA association that occurs as salt concentration increases originates from weaker DBP-DNA attraction and stronger screening. The effect of crowding on $\tau^{\mathrm{Encounter}}$ is simpler than its effect on τ^{Target}. Figure 4.3B shows that, regardless of salt concentration, the average time to achieve nonspecific protein-DNA association increases as the volume fraction increases. $\tau^{\mathrm{Encounter}}$ is particularly long for high values of φ due to the strong caging effect on the DBP, which restricts its 3D diffusion. The longer $\tau^{\mathrm{Encounter}}$ at high volume fractions due to confinement by the crowders is reflected in the lower D_3 value as φ increases (Fig. 4.3D). Finally, we measured τ^{Target} after initially placing the DBP far from the DNA. The results (Figure A.2) show that the time needed to find the target site can be governed by the effect of crowding on both 1D and 3D diffusion. Figure A.2 shows that the total time resembles that of τ^{Target} when placing the protein on the DNA, which suggests that DNA search under crowded conditions is dominated by the timescale of linear diffusion rather than 3D diffusion.

4.2.3 The effect of the mass of the crowders on the kinetics of the facilitated diffusion

To understand better the effect of crowding, we investigated how the mass of the crowders may influence their ability to modulate the kinetics of searching the DNA. Up to this point, the presented results were obtained using crowder molecules with a constant small mass of

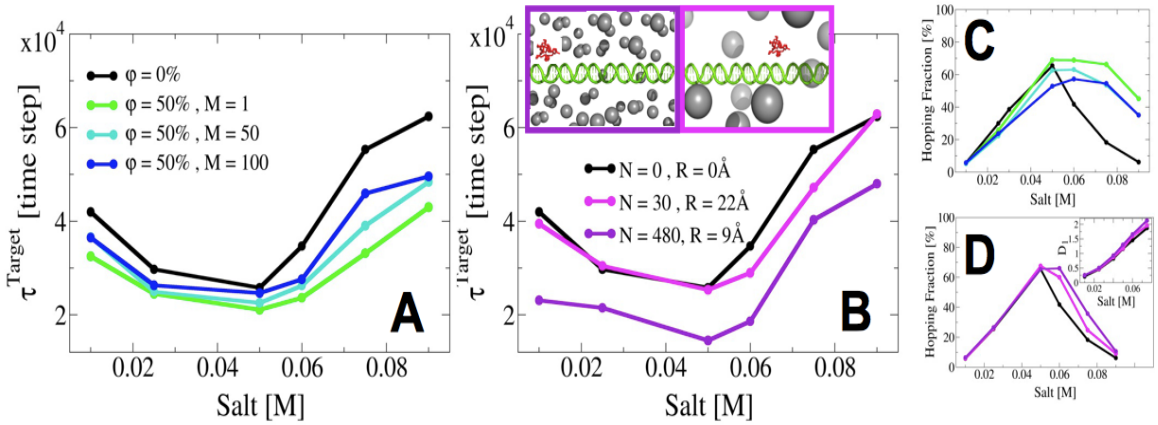

Figure 4.4 – The effect of the mass and dimension of the crowding particle on DNA search. A) τ^{Target} as a function of salt concentration at a crowder fractional volume (φ) of 0% or 50% for crowding particles of different masses (and R = 9Å). B) The τ^{Target} adopted by a DBP at a φ of 12.5% at different salt concentrations using crowding particles of R=0 Å (no crowding particles), R=45 Å or R=18 Å (M = 1). The snapshots illustrate the two conditions of different N and R to achieve the same fractional volume of 12.5%. The DBP is shown in red, the crowder molecules as grey spheres, and the DNA molecule appears in green. C) Effect of fractional volume on the proportion of hopping utilized by a DBP on DNA for crowding particles with the various masses examined in A). D) The effect of particle size on the usage of hopping as a DNA search mechanism for the same particle sizes as examined in B). The inset shows the diffusion coefficient D_1 as a function of salt concentration during simulated protein scanning of the DNA.

M=1. The mass of each amino acid bead of the DBP was also 1, such that the DBP was 93 times heavier than a crowding molecule. In order to examine mass effects, we repeated the calculations for larger crowder particles with mass of either M=50 or M=100.

In Figure 4.4A, we present the effect of mass on τ^{Target} by comparing the value of the latter at φ =0% and at φ =50% for particles of mass M=1, 50, or 100 at various salt concentrations. As was discussed above (Fig. 4.3A), the presence of crowding particles facilitates the kinetics of the system by restricting the protein?s movement so that it remains closer to the DNA axis, so enabling it to explore new sites. With increasing mass, τ^{Target} values gradually increase, until they approach the kinetics of the dynamics in the absence of crowding (φ =0%). In addition, Figure 4.4C shows the usage of the hopping search mode in the presence of different crowding masses. Similarly to what was seen in Figure 4.4A, increasing the mass of the particles also diminishes the strong effect of crowding on the proportion of hopping used. The effects of an increase in mass mimic those due to crowding, as was seen in Figure 4.2A.

A possible explanation for these effects is that introducing slowly moving crowding particles increases the viscosity of the intracellular environment, and therefore particle-DBP collisions become less frequent. Heavier crowding particles affect the linear dynamics of the proteins along the DNA differently at different salt concentrations. At low salt concentrations, fewer collisions between higher-mass particles and the DBP (blue line in Fig. 4.4A)

results in a smaller linear diffusion coefficient and therefore a slower search (larger τ^{Target}) than that associated with lower-mass particles (green lines in Fig. 4.4A), yet it is faster than in the case of no crowding (black line, Fig. 4.4A). The introduction of higher-mass particles also results is an increase in $\tau^{\mathrm{Encounter}}$ (Fig. A.3), further exemplifying the caging effect of the crowders. At higher salt concentrations, the mass of the crowding particles affects the hopping mode, with the presence of heavier crowding particles causing less hopping and thus slower kinetics for the DBP to find its target site (Fig. 4.4C).

4.2.4 The effect of the size of the crowders on the kinetics of facilitated diffusion

In addition to the mass of the crowding particles, their dimension (i.e., their radius R) may also affect their overall effect on diffusion along DNA. Indeed, the effects of molecular crowding on biochemical reactions are known to be dependent on the reactants? size and geometry (8). Here, the crowding particles are assumed to be spherical and their dimension can be tuned simply by changing their radii. Specifically, large particles experience entropic attraction forces in the presence of smaller particles (132), thus smaller particles are expected to have a stronger effect on binding. We began by studying the kinetics effect as a function of salt concentration by measuring τ^{Target} with crowding particles that were either smaller or larger (Fig. 4.4B) than the radius of the DBP. A stronger excluded volume effect and thus more pronounced crowding effect is expected for larger particles, when keeping the concentration of the crowding particles constant. To avoid this trivial scenario, we compared two systems with different particle radii (R=45 Å or R=18 Å) but the same fractional volume, which we achieved by changing the number of particles, N (Fig. 4.4B). We examined the effect of particle size for (φ=12.5% and compared it to a system with no crowding particles, ($\varphi = 0\%$. Decreasing the size of the crowding particles (while increasing the number of crowding particles in order to keep (φ constant) results in a larger effect in which τ^{Target} is smaller than for larger crowding particles. When the crowding particles are smaller, the DBP encountered a larger number of particles, which results in a strong decrease in the values of τ^{Target} compared with the values of the larger particle at all salt concentrations (Figure 4.4B). As can be seen in Figure 4.4D, this increase in search efficiency in the presence of crowding particles having smaller dimensions is also accompanied by greater utilization of hopping at higher salt concentrations than is seen in the presence larger crowding particles, and increases in in the values of the D_1 coefficient (see inset of Figure 4.4D).

4.2.5 Monte-Carlo simulations of DNA search in the presence of molecular crowding

We supplement the CGMD simulations with MC simulations, as the latter are valuable in evaluating the effect of modifying some key parameters of the facilitated diffusion mechanism on the kinetics. In the Monte-Carlo simulations, the protein and the crowding particles

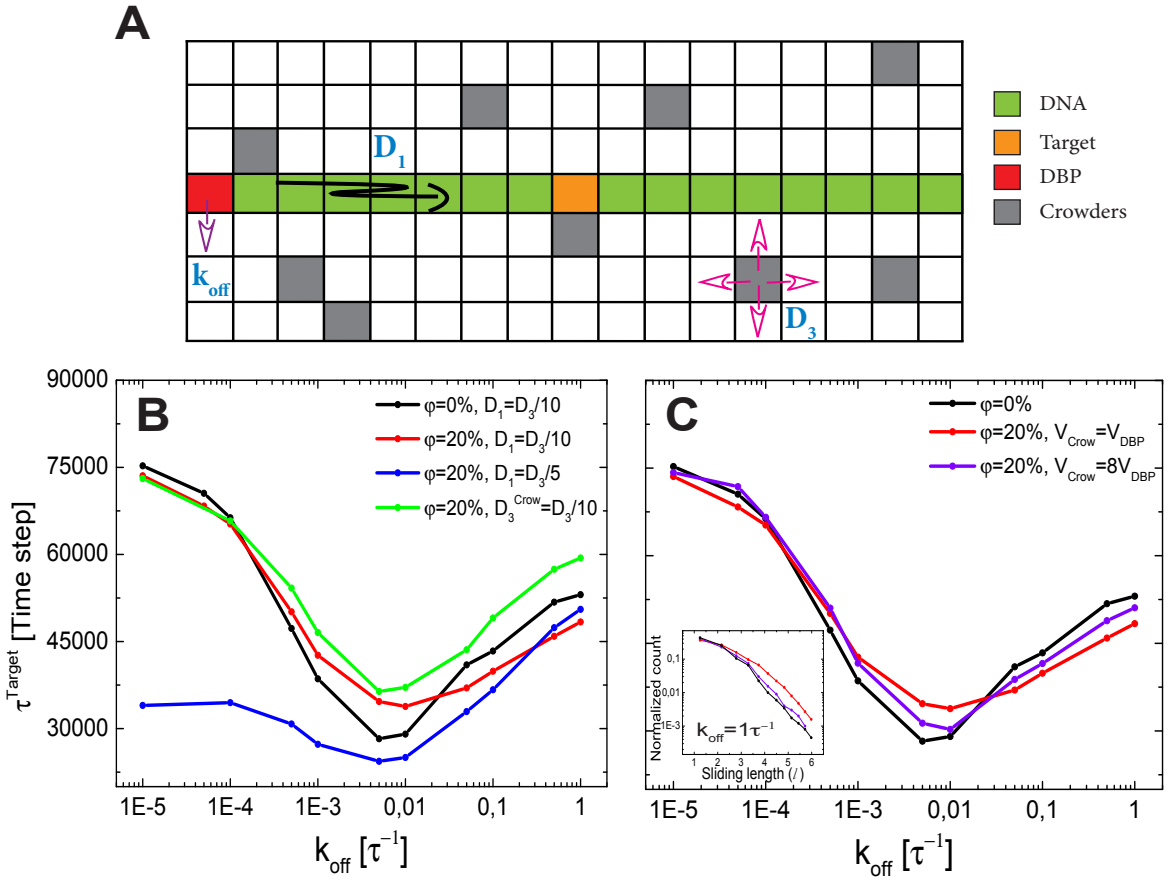

Figure 4.5 – On-Lattice MC model for target finding. A) Schematic view of the model. A DBP, in red, diffuses by hopping either to one of the six nearest neighbor lattice sites, if it is in the bulk solution, or to one of the two neighbor lattice sites, if it is sliding along the DNA, which is represented in green. Crowders represented in gray diffuse only in the bulk solution and are not allowed to bind the DNA. Molecules have square geometry, and diffusion is only allowed if the destination site is unoccupied. Note that the simulations use a 3D lattice rather than the 2D one depicted schematically here. B) Average target finding time as a function of k_{off} for different volume occupation fractions φ, 1D diffusion constants D_1 and crowder masses. C) Average finding time as a function of k_{off} for different crowder sizes. Inset) Sliding length distributions for different crowder sizes.

move on a lattice and their dynamics is dictated by the D_1 and D_3 parameters and the salt effect is mimicked via the k_{off} parameter (see Fig. 4.5A). Figure 4.5B shows the average time to find the target site as a function of the unbinding rate k_{off} for different levels of molecular crowding. When no crowders are present, the average finding time follows a non-monotonic behavior with a minimum at an optimal value of k_{off}, similarly to the CGMD results shown in Fig. 4.3A. As molecular crowding increases, the sliding length increases because of less dissociation from the DNA (Fig. A.4). This effect is mostly beneficial for large values of k_{off}, but for small values of k_{off} (i.e., low salt concentration), the presence of crowders does not affect the search kinetics. A fundamental difference between the CGMD and the MC approaches is that, while in the former the D_1 increases with salt concentration, in the MC D_1 is constant and is decoupled from k_{off}. To examine the effect of D_1 on the search kinetics, we simulated the search under crowded conditions and utilizing higher values for D_1. Figure 4.5B shows that lower values of k_{off} paired with higher values of D_1 results in a shorter τ^{Target}, however large values of k_{off} have no major effect on search kinetics. The faster search under conditions of low k_{off} and higher D_1 values is similar to the effect reported in Fig. 4.3A, in which crowding produces a shorter τ^{Target} at low salt concentrations.

To implement the different mass of the DBP and the crowding particles in the MC simulations model, we modified the crowder's diffusion constant D_3^{Crow}. Figure 4.5B shows the average finding time as a function of k_{off} for crowding particles that are heavier and diffuse with D_3^{Crow}=D_3/10. Increasing the mass of the crowding particles affects the search kinetics only at high values of k_{off} under which the kinetics is dominated by D_3. When the DBP and the crowders have the same diffusion constant, the excluded volume enhances target finding. Nevertheless, when crowders diffuse more slowly than the DBP, their effect on target localization is negative, suggesting that slow crowders affect the 3D diffusion of the DBP more strongly than crowders with the same diffusivity. This behavior is similar to the results from the CGMD simulations (Fig. 4.4A). We note that, at the limit of very slow crowders, percolation effects can take place, with the result that the DBP would never find its target.

Next, we used the MC model to study crowders that are bigger than the DBP. Fig. 4.5C shows the average finding time as a function of k_{off} for two cases in which the volume fractions are the same but the particles have different sizes: A larger number of smaller particles is seen to result in faster kinetics. The origin of this enhanced kinetics is more effective excluded volume effects, see inset Fig. 4.5C. We point out that a similar effect is reported for the CGMD simulations (Fig. 4.4B), however the effect is not identical, most likely because the CGMD also involves a change in the D_1 value that was not incorporated into the MC simulations, which focused on the effect of the dimensions of the crowding particles. The MC simulations support the CGMD observations that molecular crowding increases the linear diffusion coefficient D_1, that the presence of lighter crowding particles can speed up the search (by increasing the D_3

value), and that smaller crowding particles may be more effective in enhancing facilitated diffusion.

4.3 Conclusions

Recent studies contributed to understanding of various facets of the mechanism of DNA recognition by proteins. In addition to the typical complexities of biomolecular recognition, which demand structural and chemical complementarity, the interactions between protein and DNA require an extensive search that is further complicated by conditions in the cellular milieu. The *in vivo* aspects of protein-DNA recognition are just starting to be addressed, but some studies have surprisingly suggested that the complexity may not significantly exceed the *in vitro* scenario. Actually, it was indicated that the crowded cellular environment and particularly obstacles on the DNA may not necessarily impede the search kinetics (28; 33).

In this paper, using CGMD and Monte-Carlo simulations, we demonstrated how molecular crowding influences DBP dynamics as it searches DNA. We have shown that DBP search efficiency in the presence of crowding may be improved as its dynamics becomes confined to the DNA regions between the crowding particles. This finding may imply that the excluded volume effects of crowding particles can sequester sampling to within certain DNA regions while reducing the possibility of the DBP escaping the DNA. The sequestration effect of crowding may result in a faster search when the target site is located in the accessible region along the DNA. We demonstrate how search efficiency is altered under varying volume occupancies and show that this effect strongly depends on salt concentration. We suggest that, at increasing occupancies, crowding has an influence on the partitioning between the searching modes (sliding, hopping, and 3D diffusion) adopted by the DBP, in a similar manner to the effect of decreasing salt conditions.

Although crowding particles may restrict the DNA region that is accessible for searching, their increasing number could alter the nature of hopping events, and so influence the value of the D_1 coefficient. This influence becomes considerable at higher salt concentrations, at which the protein performs more hopping at the expense of 3D diffusion. This observation serves as another example of the importance of hopping as a DNA search mode. The enhanced hopping dynamics may also be viewed as an outcome of frequent collisions between the DBP and crowding particles, which increase as the fractional volume of the latter increases. The mass of the particles, their size, and the temperature of the system are also important parameters influencing the characteristics of linear diffusion, in addition to their entropic effect via the excluded volume interactions. Decreasing the mass of the crowding particles, decreasing the dimension of the crowding particles, decreasing their number, or using higher system temperatures increases the number of particle-protein collisions. Consequently, the linear diffusion of the DBP along the DNA is characterized by higher D_1 due to faster sliding or due

to larger usage of the hopping searching mode. As a result, crowding effects become more localized and the protein may perform less efficient sampling of the linker region and increase its search time. Thus, the synergism between crowding occupancy, the properties of crowding particles, and the salt concentration may facilitate the search and contribute to the ability of proteins to navigate through the complex DNA organization to find their regulatory binding sites.

We thus provide evidence that D_1 can be affected not only by the salt concentration (which changes the hopping propensity) but also by the volume fraction, because at low salt concentrations, collisions increase the D_1 of sliding whereas higher salt concentrations increase the hopping propensity. Furthermore, D_3 is also affected by increased crowding because of the effect of confinement, which may restrict 3D diffusion as crowding increases. Thus, this study suggests that crowding may affect search kinetics not only via excluded volume effects but also by modifying both D_1 and D_3. Thus, it is evident that the volume fraction, φ, is not a sufficient parameter to capture the effect of crowding on DNA search, but rather the molecular properties of the crowding molecules (such as their mass and molecular dimension) may control their overall effect. This is in agreement with a recent study (18) which suggested that a crowded cellular environment could influence the mechanism of DNA damage recognition by an enzyme to the same extent as any property of the enzyme itself or the DNA. We speculate that, in addition to the simple excluded volume effects of inert polymers, the presence of crowding particles with an affinity to either proteins or to the DNA (18) may modify the facilitated diffusion mechanism.

4.4 Materials and Methods

The effect of molecular crowding was studied using off-lattice CGMD and on-lattice Monte-Carlo simulations.

4.4.1 Coarse-grained molecular dynamic simulation model

In CGMD simulations, the protein is represented by a single bead per residue placed at the C_α position and the protein dynamics is governed by its native-state topology (187). The DNA is modeled by three beads per nucleotide (representing phosphate, sugar and base) that are positioned at the geometric center of the represented group. We modelled the double-stranded DNA (ds-DNA) in two different ways. In the first model, given the high persistence length for DNA of about 50 nm, it was represented as a rigid and static molecule. It was modeled in its canonical B-form and was centered on and aligned with the Z-axis. To incorporate flexibility into the DNA, the second model used a three sites per nucleotide (3SPN.1) approach, which was shown to capture the thermodynamic properties of various DNA sequences under different conditions (188). The crowding macromolecules were represented by uncharged spheres

that were initially located at random positions and that occupied a total volume determined by their number (N) and their radius (R). Their velocity depends on their mass (M) (see Fig. 4.1).

We performed CGMD simulations of a human DNA-binding protein Sap-1 (PDB code: 1bc8), in the presence of 100 bp ds-DNA of poly GC. Sap-1, a 93 amino acid globular protein with a3 total of 15 and 6 positively and negatively charged residues, respectively, uses a winged helix DNA-binding domain to activate transcription. The protein was simulated by a native topology-based model that excluded nonnative interactions and used the Lennard-Jones potential to represent native contact interactions. Electrostatic interactions acting between all the charged beads in the system were modeled by the Debye-Hückel potential (155; 189).

The dynamics of the protein along the DNA was simulated using Langevin dynamics (155; 190). The simulations were performed at a constant temperature below the unfolding temperature of Sap-1 (i.e., $T_{simulation} \sim T_f$, where T_f is the equilibrium protein folding/unfolding temperature), and were analyzed in terms of sliding and hopping (together termed 1D diffusion) and 3D diffusion. All runs were simulated for 1×10^8 time steps that allow extensive DNA sampling by the protein and transitions between sliding, hopping, and 3D search modes. To differentiate between protein sliding, hopping, and free 3D diffusion, we used the definitions defined in ref. (155). 1D diffusion along the DNA was used to calculate the mean-square displacement profiles along the Z-axis (155; 179). To address the efficiency of the DNA search, we placed the protein on the ds-DNA at a fixed position near one of its edges at time zero, and measured how much time elapsed until the DBP arrived at a target site located in the middle of the 100 bp ds-DNA. The mean arrival time to this predefined target site over 100 simulations is indicated by τ^{Target}, where arriving at the target site was defined as reaching a distance threshold of 17 Å between the recognition helix of the DBP and the phosphate beads of the DNA target site. In a similar manner, we estimated the time required for a DBP that had been placed in the bulk solution at time zero to encounter any nonspecific DNA site (i.e., $\tau^{\mathrm{Encounter}}$). While τ^{Target} estimates the effect of crowding on 1D diffusion, $\tau^{\mathrm{Encounter}}$ measures the effect of crowding on 3D diffusion.

4.5 On-lattice Monte-Carlo simulations

The crowding particles are modeled on a 3D lattice in a box with volume $V = m_x \times m_y \times m_z l^3 = \Omega l^3$ (where l is the lattice spacing and the total number of lattice sites), with periodic boundary conditions (25). One DNA molecule is modeled as a linear arrangement of lattice sites along the y-axis. The system contains a target site and two types of particles: a DBP searching for its target site (placed in the center of the DNA) and C crowding particles that diffuse freely in non-DNA sites and occupy a volume occupation fraction $\varphi = C/\Omega$ (see Fig. 4.5A). Crowders are not allowed to bind the DNA at any time during simulations. The dynamics of

the crowders and the DBP in the bulk solution is implemented as follows: At each simulation time step (of duration τ), C randomly chosen crowders and the DBP move to one of their six neighboring lattice sites with a probability of 1/6 (reduced in the case of crowders with larger mass, see Supporting Information). Thus, crowders and the DBP diffuse with a 3D diffusion constant $D_3 = l^2/(6\tau)$. If the chosen lattice site is free, the move is accepted, but if it is already occupied, the move is rejected. We thus consider excluded volume, or steric repulsion, as the only interactions in our lattice implementation. If one of the chosen neighbor lattice sites is part of the DNA, the move of the DBP onto the DNA is always accepted. Once the DBP is non-specifically bound, one of two events can take place: It can unbind from the DNA with a rate $k_{\text{off}} = P_{\text{off}}\tau^{-1}$, where P_{off} is the DBP unbinding probability. Alternatively, the DBP can diffuse along the DNA with a 1D diffusion constant D_1. Here, unless otherwise stated, we consider $D_1 = D_3/10$, since it has been shown experimentally that, in general, $D_3 > D_1$ (32). High values of k_{off} mimic screening at a high salt concentration. The DBP can find the target site through 3D diffusion if it directly hops from the solution to the target or by sliding events along the DNA. After the DBP has found the target, the simulation is stopped. Kinetic data are shown for 2000 binding events.

Chapter 5

Effects of non-spherical crowders on protein folding

Abstract

The interior of a cell is a highly packed environment that can be occupied up to 40% by different macromolecules. Such crowded media influences different biochemical processes like protein folding, enzymatic activity and gene regulation. In this work, we study protein stability under the presence of crowding agents that interact with the protein by excluded volume interactions. In general, presence of crowding agents in the solution enhances the stability of the protein native state. However, we find that the effects of excluded volume not only depend on crowding occupancy, but also on the crowder's geometry and size. Specifically, we find that polymer-like crowders have stronger influence than spherical crowders, and that this effect increases with polymer length, while it decreases with increasing size of spherical crowders.

5.1 Introduction

Protein folding is at the heart of cell activity, since in order to be active and perform their functions, proteins need a defined compact 3-dimensional configuration, the so called native structure (97; 98; 100). Thus, proteins are not random heteropolymers, but complex molecules that organize themselves based on their amino acid sequence (99; 191). How proteins fold into the native structure has been the center of many theoretical, computational and experimental studies (10; 95; 96; 97; 98; 192; 193). The most accepted model for protein folding describes the protein folding energy landscape as a rugged funnel, in which the native structure is at its global minimum (95; 194; 195). In this scheme, a coiled chain of amino acids is directed towards the interior of the funnel as native contacts are formed. Since these native contacts stabilize the protein, protein folding has a preferred direction of flow. It is believed that such interactions were achieved through evolution by choosing the sequences

that lead to those attractive native interactions. It is then said, that proteins follow the so called principle on minimal frustration (95; 113).

Under physiological conditions, protein folding takes place in a crowded cytoplasm rather than in a dilute solution. In particular, under fast growing conditions, the *Escherichia coli* cytoplasm can be occupied up to 40% by macromolecules, such as the DNA, RNA molecules, proteins and sugars (2; 6; 34; 87). Accordingly, it is expected that protein folding under *in vivo* conditions differ from protein folding in dilute solutions.

Crowded conditions can not be avoided *in vivo*. Thus, a proper analysis of how proteins fold under physiological conditions needs to consider the effects that excluded volume interactions have on the folding process. In particular, these effects can be classified in two principal kinds: Effects on binding equilibria, and effects on diffusion and reaction rates (6; 7; 8). Molecular crowding has a positive effect on the binding equilibrium and stabilizes the bound state (7; 8; 25; 124). On the contrary, effects of crowding on diffusion are negative, as excluded volume hinders molecular motion in the solution (22; 23; 24). Thus, effects of molecular crowding on protein folding can be difficult to predict. These effects have been the focus of experimental, computational and theoretical studies (4; 7; 9; 10; 11; 16; 17; 19; 117; 118; 119). The results of these studies indicate that molecular crowding promotes stabilization of the native state by increasing the free energy of the unfolded configurations (4; 10; 11). This is because the cost of inserting a larger unfolded configuration into a hard-sphere fluid is greater than inserting a smaller folded configuration (10; 11). It has also been shown that crowder size influences the strength of the effects of excluded volume on protein stability. Specifically, protein stabilization increases as the size of the crowders decreases (11).

In this study we address the effects of molecular crowding on protein folding by using a structure-based model (SBM) of the protein HigA. SBM are based on two main theoretical foundations. One is the principle of minimal frustration. The second one is the hallmark of SBM, is the funnel-like folding energy landscape (113; 196). Thereby, SBM reduce drastically the complexity of the protein folding force field, and time scales of folding can be achieved. SBM have been developed for both full-atom (197), and coarse-grained representations (114; 115). Moreover, SBM have been successfully used in the modeling of protein folding (113), RNA folding (198; 199) and multimeric folding and binding (200).

We provide new understanding of the effects that macromolecules with different shapes, geometries and sizes, have on protein folding and protein stability. Specifically, we consider the effects that spherical crowders with different sizes and two types of polymer-like crowders with different lengths have on protein folding.

5.2 Methods

5.2.1 Coarse-grained model

We implement a coarse-grained model in which the protein is represented by a single bead per residue centered at the C_α position. We run simulations for the HigA protein (PDB code 2ICT), the bacterial antitoxin HigA from *Escherichia coli*. This is an 89-amino-acids-long non-globular protein, that does not have any β structure (201), see Fig. 5.1. The protein was implemented using the structure-based modeling software eSBMTools, which are python tools that interface with GROMACS (115), see Fig. 5.1.

Simulations are run for a total time of 250 nsec. At every 5 pico sec, a snapshot of the protein structure is taken in such a way that a total of 50,000 protein structures are obtained at the end of each simulation. The statistics are done with these sets of data. We point out that due to coarse-graining of the model, one cannot easily convert the time steps to real time. However, our implementation accurately reproduces the unfolding-folding dynamics of the protein.

We use periodic boundary conditions that take into account interactions between images of the protein and the crowding agents with themselves. The dimension of the simulation box is given by the implementation of the eSBMTools. Specifically, for all of our simulations, the box size has a total volume $V_{\rm tot} = 14.78 \times 12.17 \times 13.75$ nm^3.

In the analysis of our simulations, we use the number of native contacts Q as the protein order parameter. Hence, large Q values represent a protein close to its folded state, whereas low Q values represent a protein in an unfolded configuration. The native structure provides the maximum number of native contacts Q. In particular, for the considered HigA protein $Q_{\rm max} = 139$. For this frustrated energy landscape, energy input is required to bring the protein out of the bottom of the funnel and change the protein conformation. Thus, we change the conformations of the protein by modulating the temperature (in reduced GROMACS units). For all considered cases, the minimum considered temperature is $T = 50$. Here, the protein is folded and the average Q value is $\sim Q_{\rm max}$. We increase the temperature until $T = 200$. At this high temperature, the energy input is high enough to take the protein out of its native state. Thus, with this temperature, we ensure that the protein is present in an unfolded state.

5.2.2 Crowding agents

The aim of our study is the characterization of the effects that molecular crowding has on protein folding and protein stability. We thus implement crowders that occupy a volume fraction ϕ. A natural starting point is to implement crowders with spherical geometry and with the same size of the C_α of the coarse-grained protein model. Crowders are manually created, and they are set to interact with the protein only via steric interactions. The eSBMTools imple-

mentation provides the potential parameters of the protein, including the $C_{12} = 1.67 \times 10^{-5}$ value of repulsion term of the Lennard-Jones potential, $V_{LJ}(r) = C_{12}/r^{12}$. However, the radius of the C_α is not provided. This potential does not directly define the radius of the C_α beads und hence the volume fraction occupied by crowders. Therefore, we determine the radial distribution function from simulations with 1000 crowders mplemented with the same C_{12} value, and for three temperatures $T = 50$, 120 and 150. Next, we obtain the radial distribution function (RDF) of the system. For all three temperatures, the obtained C_α radius is ~ 0.33nm. Thus, each C_α has a volume $V_{\mathrm{C}_\alpha} = 0.15\mathrm{nm}^3$, and occupies a volume fraction $\phi_{\mathrm{C}_\alpha} = V_{\mathrm{C}_\alpha}/V_{\mathrm{tot}} = 6.06 \times 10^{-6}$.

Crowders with larger volumes are implemented after proceeding in the same way as described above. For a given value of the parameter C_{12}, the RDF is obtained together with the radius of the crowding agents. Then, the volume of the spherical crowders is obtained, and the box can be filled with a specific number of crowders that occupy the desired volume fraction ϕ.

We continue by implementing crowders that have different geometries. Here, we consider polymer-like crowders that are implemented as linear arrangements of beads. Each bead is set to have the same volume as the C_α. The polymer-like crowder does not have any backbone interactions, except for a bond-like potential that keeps the polymer chain together. Different polymer lengths are achieved by adding C_α beads, each with their corresponding bond potential, to the chain.

5.3 Results

Throughout this study we use the small protein HigA as a model protein. This protein is investigated with molecular dynamics simulations using a coarse grained structure based model that describes the protein as a chain of beads that are located at the positions of the C_α atoms with nonlocal interactions defined by the native topology of the protein. In these simulations, the HigA protein displays two well-separated states, folded and unfolded. The histograms of the number of native contacts, Q, is monomodal at high and low temperatures, with typical Q values > 120 and < 40, respectively, but bimodal for temperatures around $T = 109$ (Fig. 5.1). To quantify folding, we determine the following three parameters: the average number of native contacts, $\langle Q \rangle$, the coefficient of variation $CV = \sigma_Q/\langle Q \rangle$, where σ_Q is the standard deviation of the Q values, and the radius of gyration of the protein, R_G. The three parameters are plotted as functions of the temperature in Fig. 5.2A-I. The average number of native contacts decreases steeply in a narrow range of temperatures around the folding temperature (Fig. 5.2A, D and G). In that temperature range, the fluctuations of Q are increased, and the folding temperature can be precisely defined by the position of the maximum in the coefficient of variation, which is found here to be $T \simeq 109$ for the free scenario

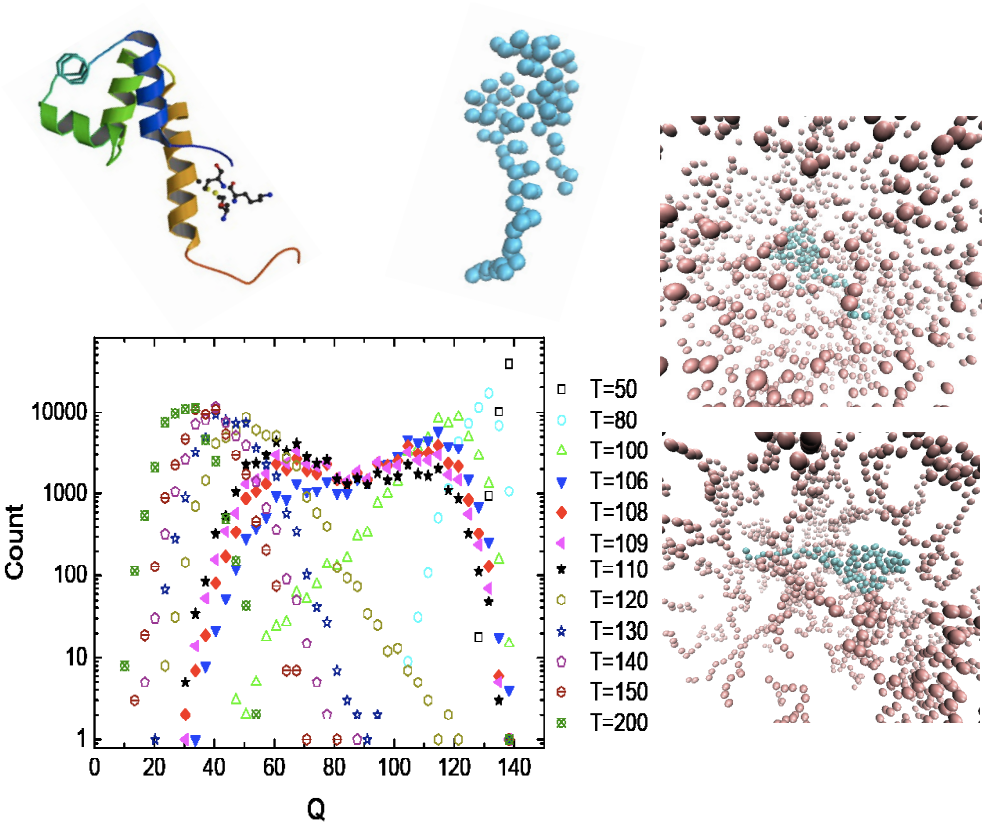

Figure 5.1 – HigA protein and its coarse-grained representation given by eSBMTools. Histograms of the number of native contacts Q. As the temperature increases, the protein is destabilized, and the distributions shift towards low Q values. Snapshots of the system with spherical (top) and polymer-like crowders (bottom) at $\phi = 0.1$.

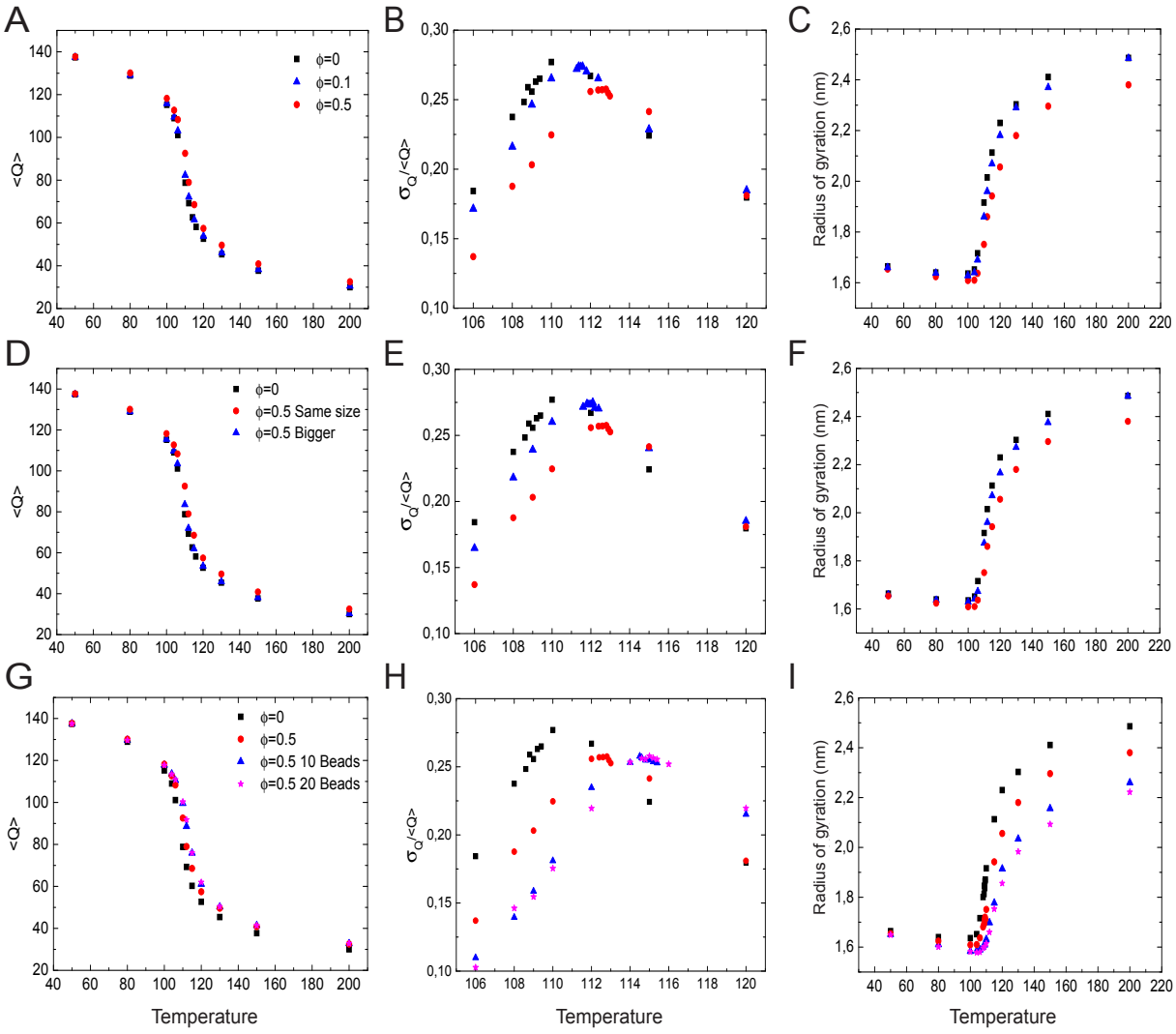

Figure 5.2 – Effects of crowders on protein folding. A-D-G) $\langle Q \rangle$ as a function of the temperature. Presence of crowders shift the data towards higher temperatures. Molecular crowding enhances protein stability. Polymer-like crowders have the strongest effects among all considered crowders. B-E-H) CV as a function of the temperature. The folding temperature, which is obtained as the maximum of CV is shifted to higher values. Among all considered crowders, polymer-like crowders have the strongest influence on protein stability. The shift of the maximum in CV depends on the length of polymer crowders. C-F-I) Radius of gyrations as a function of the temperature. Macromolecular crowding enhances compactness of the protein. The effect on T_{F} correlates with the influence of crowders on the radius of gyration.

(Fig. 5.1B black squares). The loss of the native structure as measured by Q is accompanied by loss of the compactness of the structure, which is quantified by the radius of gyration (Fig. 5.2C, F and I).

To study the effect of crowders on the folding of our model protein, we include different types of crowders in the simulations. All crowders are inert and interact with the protein only through volume exclusion. Specifically, we consider spherical crowders of the same size as the C_α beads, larger spherical crowders and polymeric crowders consisting of chains of beads with the same sizes as the C_α beads of the protein.

Fig. 5.2A-C shows the temperature dependence of the three parameters quantifying folding for different volume fractions of the small spherical crowders. The presence of crowders shifts the decrease in the average number of native contacts to higher temperatures, indicating that crowders stabilize the folded state, in agreement with earlier experimental studies (117; 119; 202; 203). Consistent with that observation, the peak in the CV and the steep increase in the radius of gyration are also shifted to higher temperatures. The stabilization of the protein can be quantified by the folding temperature, which increases with increasing volume fraction of crowders (Fig. 5.3A).

Larger crowders have qualitatively the same effect, but the shift of the folding temperature is smaller than for the small crowders as depicted in Fig. 5.2C-E. Indeed when keeping the occupied volume fraction constant, the folding temperature is seen to decrease as a function of the crowder volume, see Fig. 5.3B. That smaller crowders have a larger effect is consistent with previous studies (11; 204; 205). One reason for this observation is that depletion forces are stronger for smaller crowders (35; 206).

Polymeric crowders result in a pronounced shift of the folding temperature. At the same volume fraction, their effect is considerably stronger than the effect of spherical crowders. As for spherical crowders, the effect depends on the crowder size. Surprisingly, the size dependence is not only more pronounced than for spherical crowders, but the shift of the folding temperature has the opposite direction: While the folding temperature decreases with increasing size of spherical crowders, it increases with increasing length of the polymeric crowders. Thus, the effect of molecular crowding on protein folding depends on the volume fraction, the size and the geometry of the crowders.

Finally, we consider the HigA protein as a self-crowder by observing the folding of a 'labeled' protein in a dense solution of unlabeled proteins. In this case, the volume fraction occupied by crowders is determined from the average radius of gyration of the protein at the fording temperature T_F. As a consequence, fewer crowders are needed as the temperature increases, because of the increase of the radius of gyration (Fig. 5.2C black squares). In that case, the folding temperature decreases to $T_\mathrm{F} = 111.9$.

We then ask weather the internal interactions that provide crowders with native structure have an effect on the stability of the protein HigA. Therefore, we run simulations in which we

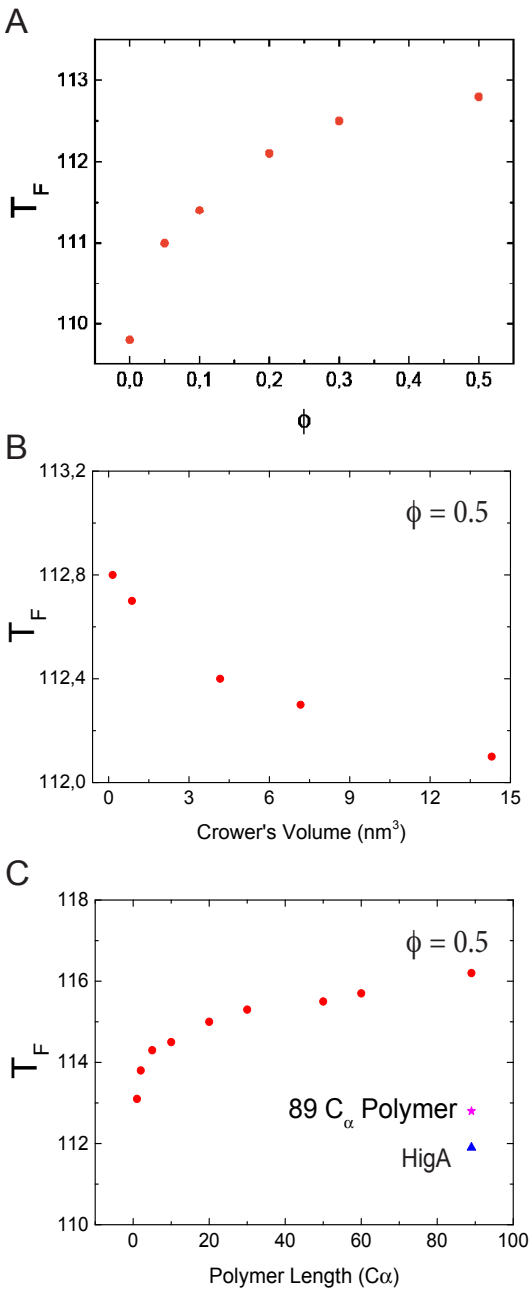

Figure 5.3 – Effects on the stabilization of the HigA protein by crowding agents. A) T_{F} as a function of ϕ, for crowders with the same size as the C_α atom. T_{F} increases together with ϕ. B) T_{F} as a function of the spherical crowder's volume for $\phi = 0.5$. Stabilization of the protein diminishes for larger crowders. To keep a constant volume occupation fraction, the number of crowders need to be adjusted. C) T_{F} as a function of the polymer length. T_{F} increases as the crowders become longer (red points). At $\phi = 0.5$, 89-beads long polymers shift T_{F} in more than 6 temperature units. The blue triangle represents the case in which the crowding agents are HigA proteins. Here, $\phi = 0.5$ was calculated by taking the average radius of gyration of the protein as the geometrical parameter. T_{F} goes down to $T_{\mathrm{F}} = 111.9$. Interestingly, after deletion of all backbone interactions of the HigA crowders (except bond ones), T_{F} increases by a temperature unit (pink star).

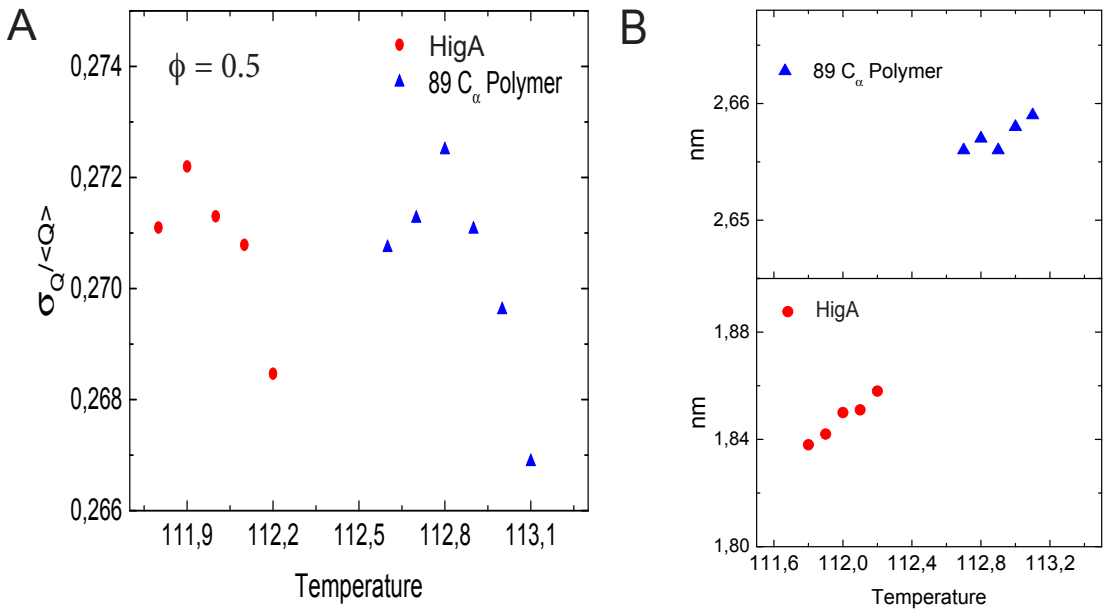

Figure 5.4 – Effects of polymer-like crowders with and without defined conformation on protein stability. A) At $\phi = 0.5$, T_F is larger when polymer-like crowders without backbone interactions are present in the simulation box. B) Radius of gyration of the two kinds of crowders. Crowders without a native structure (top) are more extended in space than crowders with a native conformation (bottom).

keep the number of crowder beads, but remove all the internal interactions within the crowders. It is, crowders are 89-long bead polymer chains with only bond interactions. Notably, deletion of interactions within the polymer increase protein stability by about 1 unit of temperature, see Fig. 5.3D. This suggest that long crowders without any specific structure stabilize more the native state of the HigA protein. In this study, we only consider the unavoidable steric effects of molecular crowding, which are of entropic nature (4; 8; 173; 207; 208). Since polymeric crowders without internal interactions can acquire more configurations than protein crowders, the entropic contribution for the former scenario is larger than for the latter. In Fig. 5.4A, we present the CV for both types of crowders, in which the shift in the folding temperature can be observed. In Fig. 5.4B, we show the average radius of gyration for both types of crowders. We see that polymeric crowders (89 C_α long) without a native structure are more extended than protein crowders that are more compact because of backbone interactions.

5.4 Concluding remarks

In this paper we used a structure-based model implementation of the protein HigA to study its stability and folding under molecular crowding conditions. We implemented spherical crowders to revisit protein stability. As the volume occupation fraction increases, the folding temperature of the protein increases and the native structure of the protein is preserved for higher temperatures. This effect depends on the size of the crowders. At a given volume oc-

cupation fraction, crowders with big radii exert weaker effects on the stability of the protein. We showed that the folding temperature decays as a function of the crowder's volume.

In addition, we implemented two types of polymer-like crowders, one type without and another one with a defined native structure, i. e. a protein. We showed that protein stability and the radius of gyration are more influenced by polymeric crowders (without backbone interactions) than by spherical crowders. This effect becomes larger as the length of the polymeric crowders increases. When comparing the effects of crowding on protein stability between the two types of polymer-like crowders, we found that polymeric crowders without a native structure promote protein stability more than protein crowders.

Effects of molecular crowding on protein folding can be difficult to predict, because of the many variables that need to be taken into account. Here, we have shown that size, geometry and local organization of crowding agents lead to different levels of protein stability. Although experimental procedures can be challenging due the difficult task of finding 'good' crowding agents, it can very useful to perform detailed studies on the stability of proteins, in which crowding agents with different geometries and size are used.

During this study, we have discussed the repulsive interactions of crowding agents. However, crowders can also interact with the protein via attractive interactions. It has been shown that typically, attractive interactions lead to destabilization of protein complexes and protein folding (173; 207; 209; 210). While the effect of repulsive interactions are entropic and depend on the excluded volume, attractive interactions are enthalpic and depend on the contact area between the protein and the crowder (173). These interactions will be considered in future work, but we already anticipate that their influence on protein folding might also depend on the geometry of the crowders. Specifically, long polymer crowders can have larger contact areas than spherical crowders or crowders with a given structure. Thus, the positive influence on the folding temperature presented in this work can be reduced or even become become negative if the interaction energy is strong enough.

5.5 Acknowledgements

We are grateful to Carlos Navarro for helping us to set up the eSBMTools implementation. We thank Alexander Schug and Claude Sinner for useful discussions and orientations.

Chapter 6

Summary and Outlook

6.1 Overview of the main results

The interior of a living cell is a highly crowded environment, with macromolecules occupying up to 40% of the total cell volume (1; 2). Molecular crowding has thermodynamic effects on binding equilibria, while also decreasing the encounter rate between reactants by affecting diffusion constants (6; 8; 22; 23; 24). Therefore, biochemical reactions that occur in such crowded environments differ from reactions that take place under test tube conditions. A great effort has been made to characterize and understand the effects that molecular crowding has in the contexts of protein-protein binding, gene expression, enzymatic activity and protein folding (4; 7; 8; 9; 10; 11; 13; 14; 15; 16; 17; 18; 19; 20; 21)

In this work, we focused on the effects that molecular crowding and steric interactions exert on three different biochemical reaction processes: First, we studied enzymatic reactions. Second, we studied the consequences of molecular crowding in the context of gene expression, and specifically, how a DNA binding protein (DBP) finds a specific DNA sequence. Finally, we studied the protein folding process under crowding conditions with crowders of varying size and shape.

6.1.1 Enzymatic reactions

Experimentally, it is often difficult to decouple the different effects of crowding on biochemical reactions from one another, since changing crowding conditions affects both reaction thermodynamics and kinetics. To systematically investigate these effects, we proposed and analyzed two reactions that (at equilibrium) uncouple the effects of molecular crowding on binding equilibria from the effects on diffusion. In order to do so, we ran Monte Carlo simulations on a three-dimensional lattice with volume V, in which we placed ligands, crowders and a receptor that interacted via steric interactions.

In the past, most studies have focused on the entropic effects molecular crowding exerts on protein conformation and protein complexes. Several theoretical models based on excluded volume have been proposed to account for these effects on protein association (4; 7; 8), and

it has been shown experimentally that volume occupation leads to stabilization effects of the native structure of proteins and multi-protein complexes (4; 7; 8; 9; 10; 11; 12; 37; 38; 42; 173). For protein-protein interactions, the effect is attributed to a decrease of the complex dissociation constant as volume occupation increases (7; 8; 11; 12). In Chapter 2, we showed for the first proposed reaction that, at equilibrium, this influence of crowding holds. We found that as volume occupation increases, the average binding rate remains constant, whereas the average unbinding rate decreases. We showed that this effect becomes stronger as the size of the crowders decreases in agreement with scaled particle theory (SPT). SPT provides an analytical expression for the work needed to place a spherical particle of radius R into a hard-sphere fluid (8; 35), and give a description of the nature of depletion forces (35; 206). When carefully looking at the kinetics of this reaction, we found that the average binding rate remains constant due to the contribution of two opposite effects. On one hand, as crowding increases, fast rebinding is enhanced due to the presence of crowding agents and the close proximity of the ligand to the receptor. On the other hand, if the ligand escapes form the vicinity of the receptor, the rebinding is hindered by the presence of crowding agents in the bulk solution, see Fig. 2.3. The latter effect is studied by a second proposed reaction that monitors the effects of crowders on diffusion. We showed that the presence of crowding agents negatively influences diffusion, and thus, the time needed for binding.

Even using our coarse-grained lattice implementation, we were able to obtain valuable information on the kinetics of protein-protein interactions that had not been previously observed. While the effects of molecular crowding on kinetics can be difficult to verify experimentally, more detailed computational implementations such as the ones used by Mittal *et al.* (11; 12) can help us gain a better understanding of the microscopic dynamics of protein-protein interactions. This, in turn, will allow us to interpret data that is not only at or near equilibrium.

6.1.2 Gene expression

All biological functions of DNA depend on the recognition of specific sequences by site-specific DNA-binding proteins (DBPs) (65). The binding of these proteins to their binding sites is remarkable, because proteins quickly and accurately locate small targets ($\sim 10 - 30$ bp long), in a chromosome that is over 10^6 bp long (26; 66). Specific DBP-DNA binding takes place through a facilitated diffusion process that combines 3D diffusion events in the bulk solution with 1D diffusion excursions along the DNA (69; 70). Despite different studies on this topic (32; 82; 150; 169; 174; 175; 176; 177; 170), researchers lack a detailed description and understanding of how this process takes place under crowding conditions.

We have implemented a lattice model for facilitated diffusion, in which crowding agents either in the bulk solution or bound to the DNA influence the finding of a specific DNA sequence by a DBP. In Chapter 3, we found that as the DNA is occupied by obstacles, the efficiency of

facilitated diffusion decreases in agreement with previous models (28; 88). We showed that this negative effect depends on the dynamics of crowders on the DNA. Specifically, if obstacles are tightly bound to the DNA and do not diffuse, the DBP 1D sliding lengths are truncated, and more excursions to the bulk solution are needed to overcome the presence of obstacles. We showed that for static crowders and physiological levels of DNA occupation $\sim 20\%$, the optimal average fraction of non-specific DNA binding time drops from 50% (in the absence of obstacles), to $\sim 9\%$. We found that for more dynamic obstacles, this effect is reduced in agreement with (28; 33; 74). These results provide insights into the understanding of not only steric effects on facilitated diffusion, but also the effects that different kinds of DBP have in the specific-DNA sequence dynamics. In particular, our results are in agreement with the *in vivo* experimental results obtained by Elf *et al.*, in which they modified the operators in the vicinity of the *lac* operator to investigate the effects of different obstacles on facilitated diffusion. By changing the binding sites, they found that different obstacles differently influence specific-sequence recognition (74).

We then investigated the effects that crowders in the bulk solution have on facilitated diffusion. Using coarse-grained molecular dynamic (CGMD) and Monte Carlo (MC) simulations, we showed that molecular crowding enhances non-specific DBP-DNA binding, consistent with the general results from Chapter 2. Thus, larger 1D sliding events were found when crowding agents were present. We found that at a certain set of parameters, molecular crowding enhanced specific-DNA finding, in agreement with (18). We showed that a big contribution in reducing the average finding time is that as crowding increases, the fraction of hopping events notably increases. Although this contribution could only be observed by the use of the MC implementation, the observed effect in the hopping fraction is comparable to the enhancement in the fast rebinding discussed above in the context of protein-protein interactions.

Although our two implementations of molecular crowding provide new insights in understanding effects of molecular crowding on facilitated diffusion, a complete mapping from our modeling to *in vivo* conditions is still lacking. For instance, it is not entirely clear from our results which set of conditions make facilitated diffusion under the simultaneous contributions of crowders and obstacles more efficient than the free diffusion model. In particular, our results cannot provide an explanation to the very long fractions of non-specific binding time observed in experiments ($\sim 90\%$ in contrast to the 50% and $\sim 9\%$ in the presence of obstacles) (32; 74). As a further step in modeling facilitated diffusion, DNA supercoiling effects have to be considered within crowded conditions, since their effects on facilitated diffusion have already been studied in the free scenario (77; 93; 156).

6.1.3 Protein Folding

In order to be active, proteins need to be in a 3D compact conformation known as the native state (45; 95; 96; 97; 98). Proteins are not simple random chains of polypeptides, but are complex molecules that organize themselves based on the sequences of amino acids they are made of. The relatively short timescales in which folding takes place suggest that the non-native configurations of the protein are energetically unstable and that the flow of folding is towards the native state. The fact that the native structure is unique indicates that this conformation corresponds to the global minimum of the rugged folding energy landscape (95; 101; 102). Earlier computational experiments have shown that the folding energy landscape becomes less rugged when increasing crowded conditions, enhancing destabilization of non-native configurations (10; 11). Cheung and Mittal *et al.* have shown that stabilization of the native protein structure is achieved as the media is occupied by inert spherical crowders. Thus, as volume occupation increases the folding temperatures are shifted to higher temperatures (4; 10; 11; 117). These computational predictions have been experimentally confirmed by different studies (117; 118; 119).

In Chapter 5 we investigated the influence of crowder size and geometry on protein stability. We first considered the effects of spherical crowders with a given size, and found that as volume occupation increases, so does the stability of the protein, in agreement with (10; 11; 207). Then, we varied the size of spherical crowders and found that the effect of volume occupation on protein stability is reduced as crowders increase in size. These effects of crowder size on protein stability have been previously reported in different studies (11; 12; 173). Next, we considered polymer-like crowders with and without backbone interactions. First, we showed that if we keep constant the level of volume occupation but increase the length of the polymers, the folding temperature shifts to higher temperatures and protein stability is enhanced. Thus, the size effect of polymeric crowders is opposite to the size effects of spherical crowding agents. Then, we compared the effects of two kinds of polymeric crowders with the same length: Protein-like crowders with a native state and polymer crowders without native configuration. We showed that the influence of molecular crowding is larger for crowders without backbone interactions than for crowders with a native structure. In this study, we only considered steric effects of molecular crowding. Thus, the influence of volume occupation is of entropic nature (4; 8; 173; 207; 208). Since polymeric crowders without internal interactions can visit many more configurations than a protein with a defined native state, the entropic contribution is larger for former case than for the latter, and hence, the observed effect on protein stability.

Here, we concentrated on the effects of crowding on protein folding at equilibrium. The folding-unfolding kinetics and how the folding and unfolding rates are influenced by the presence of different crowding agents will be the focus of our next study. Based on information from a study conducted by Cheung *et al.*, which showed that the folding rate non-

monotonically depends on the volume occupation of spherical crowders (10), we think it would be useful to consider the possible effects of polymeric crowders and their lengths on folding rates. We have planned a future work dedicated to this issue.

6.2 Outlook

Throughout this work we have studied different biological processes within crowded environments. Although we discussed a variety of biochemical reactions, the individual topics considered here are very broad and can be studied in more detailed. We now provide possible extension of the work we have presented in this thesis.

6.2.1 Protein synthesis as an enzymatic reaction

The methods and models we introduced in this work are quite general and can be applied to other biological processes. For instance, in Chapter 2 we considered the general scenario of an enzymatic reaction, in which a ligand reacts with its receptor in a crowded environment. The same framework can be used to model other cell processes that are not enzymatic in a strict sense. As an example, we can consider the translation of a messenger RNA (mRNA) by ribosomes. In this context, the ribosome is the receptor of a ternary complex molecule (ligand) that carries an amino acid. At a first approximation, ribosomes are translating all the time. To elongate a polypeptide chain, ribosomes read the mRNA codon by codon by sequestering the right 'charged' ternary complex. Each ternary complex is formed by the elongation factor EF-Tu, which is the most abundant protein in *E. Coli* (128), and a transfer RNA (tRNA) molecule charged with an amino acid. Since the delivery of the right charged ternary complex is diffusion limited, molecular crowding sets a minimum in the translation rate, which can negatively influence bacterial growth (127). Moreover, molecular crowding can also have effects on the 'recharging' of the sequestered ternary complex with a new amino acid, and generate depletion effects that can slow down translation. This limitation can be avoided by the presence of local charged ternary complex pools in the proximity of the ribosome, as has been suggested in (148).

6.2.2 DNA supercoiling and lattice limitations

In this work, we used lattice implementations to model different biochemical reactions. The lattice models provide a clear and simple pictures of the modeled systems, with the advantage of being computationally inexpensive and easy to implement. Along with the lattice models, we also implemented off-lattice simulations (ReaDDy and CGMD) and obtained comparable results. Nevertheless, when modeling complex biological processes, lattice models become

inefficient and difficult to implement. For example, in Chapters 3 and 4 we studied the facilitated diffusion process under crowding conditions. There, we considered the DNA as a linear template, which, in cells, is not necessarily true. It has been shown that the DNA in its supercoiled structure has implications in target finding (77; 93; 156). Although during this study other DNA geometries were implemented, we were not able to see any drastic change in facilitated diffusion. We note that complicated DNA geometries can be obtained, but they have to be manually created, a task that can be tedious and computationally impractical. Additionally, we saw in Chapter 4 that the coarse-grained molecular dynamics implementation was able to recover information that we could not with the lattice. Specifically, we showed that molecular crowding enhances hopping under certain conditions, leading to a faster finding of the target.

6.2.3 Effects of molecular crowding on folding of intrinsically disordered proteins (IDPs)

In this study, we saw that molecular crowding enhances the stability of proteins, and that this effect depends on the geometry and size of the crowders. An extension to this study is the characterization of the folding-unfolding rates of the protein. It has been previously shown that molecular crowding affects these rates in a non-monotonic fashion (10). Moreover, the effects of molecular crowding on protein folding were characterized for the protein HigA, a protein that has a defined native structure. Nonetheless, it will be of great importance to characterize the possible effects of crowding on the so-called intrinsically disordered proteins. It has been shown that such proteins (many other IDBs likely have many other roles) play an important role in the acclimatization of plants to cold weather (120; 121). Under these cold conditions, cells dehydrate and the concentration of macromolecules increases, leading to the enhancement of IDP folding.

6.2.4 Attractive effects of molecular crowding

Throughout this study, we only considered repulsive interactions that macromolecules exert via excluded volume. These interactions are generic and do not depend on molecular details, as opposed to attractive interactions of molecular crowding such as electrostatic, hydrophobic, hydrogen bonding, and van der Waals, which in fact, depend on molecular characteristics. Moreover, contrary to the stabilizing effect of entropic-excluded volume interactions on binding equilibria, attractive interactions destabilize binding due to the enthalpic penalty, which is a result of breaking favorable protein-crowder contacts when forming multicomponent complexes. Experimentally, the influence of attractive interactions has indeed shown destabilizing effects on protein folding, RNA folding and protein complexes (173; 207; 208; 209; 210; 211). Despite the advances in understanding these effects, it is still challenging

to describe the influence of attractive interactions on biochemical processes. While the entropic effects can be theoretically described by scaled particle theory (SPT) (8; 35), the enthalpic contributions are described by a more complicated thermodynamic perturbation theory approach (173; 208). Thermodynamic perturbation theory describes the effects attractive interactions have on binding free energy, to be dependent on the contact area between the protein and the crowders, as well as on the binding energy, and the geometry and size of the reactants (173; 208).

Attractive interaction can easily be included in our lattice approach. Further studies in which such interactions are implemented will be developed as a natural next step. For example, using the lattice model of enzymatic reactions, a metropolis algorithm can be implemented at every ligand-crowder encounter. This will lead to attractive interactions that can modify the effects we described in Chapter 2. We anticipate that attractive interactions between crowders and ligands will affect both binding equilibria and diffusion. Thus, in the Michaelis-Menten framework, the two negative contributions from attractive interactions can exceed the positive influence of the entropic binding enhancement, and decrease the overall synthesis rate or shift the optimal volume occupation fraction at which the reaction is most optimal.

Another biochemical process in which attractive interactions can have a strong influence is the folding of a protein. We have shown that entropic-excluded volume interactions stabilize proteins, and that this effect is modulated by crowder size and geometry. Implementation of attractive interactions, such as van der Waals interactions, between the protein and crowders can modify this stabilization picture. In this scenario, size and geometry modify the influence of the enthalpic effects, since the change in the binding free energy explicitly depends on the contact area between the protein and the crowding agents. Specifically, it will be interesting to evaluate the effects of polymer-like crowders with attractive interactions on protein folding, because this might lessen the strong effect we obtained in Chapter 5. Additionally, polymer-like crowders can explore many different configurations, so they can be in contact with more protein surface area than spherical crowders. This can lead to strong destabilization of the protein, favoring unfolded conformations. The latter effect can be modulated by changing the length of polymeric crowders.

6.3 Concluding remarks

In this study, we used computational approaches to study the effects of molecular crowding on different biochemical reactions. We showed that although the effects of molecular crowding are generic, they can be complex and difficult to predict. The extent of these effects depends on the features of the specific biochemical pathway and the kind of crowders. Thus, a general mapping of the effects of crowding to *in vivo* conditions might be difficult to achieve.

Hence, for a complete understanding of specific biochemical processes, it is necessary to account for the effects molecular crowding exerts on specific processes. Nevertheless, with the availability of new experimental techniques together with computational implementations, researchers can disentangle the role of macromolecular crowding in *in vivo* systems. Specifically, the advances in the understanding of the influence of molecular crowding presented throughout this work can be extended and applied to other biochemical pathways in order to deepen our understanding of the cell.

Appendices

Appendix A

Mechanism of facilitated diffusion during DNA search in crowded environments-SI

A.1 SI Methods

A.1.1 Search efficiency

In the Monte Carlo simulations, we reconstruct the facilitated diffusion process under different levels of volume occupation. As molecular crowding increases, molecular crowding prevents DBP unbinding and the sliding length increases (see Figure A.4A). This effect is beneficial for large values of k_{off}, see main text, but hinders target finding when k_{off} takes intermediate values. For very small values of k_{off}, the percentage of crowders do not affect target finding, because the DBP remains mostly bound to the DNA. One difference between the two used approaches (CGMD and the lattice model), is that in addition to k_{off}, the 1D diffusion constant is also modified as the salt concentration is modulated in CGMD. In the lattice model, k_{off} and D_1 are completely uncoupled. In Figure A.4B we plot the sliding length distribution for different values of k_{off} and D_1. As D_1 increases, the distributions become narrower due to the fast scanning of the DNA.

A.1.2 Mass effect

Cell cytoplasm contains a great variety of macromolecules with different geometry, size and mass (ref (6), main text). A natural case to consider is the one in which the DBP and the crowders have different mass. In the lattice model, this is implemented by changing the crowder's diffusion constant D_3^{Crow}. The average finding time follows a non-monotonic behavior, and is qualitatively similar to the case with crowders having the mass as the DBP. Interestingly, when crowders diffuse slower than the DBP, their effect on target finding is negative for large k_{off}, see Fig. 4.5B (main text). Even though slow crowders enhance DBP-DNA binding as

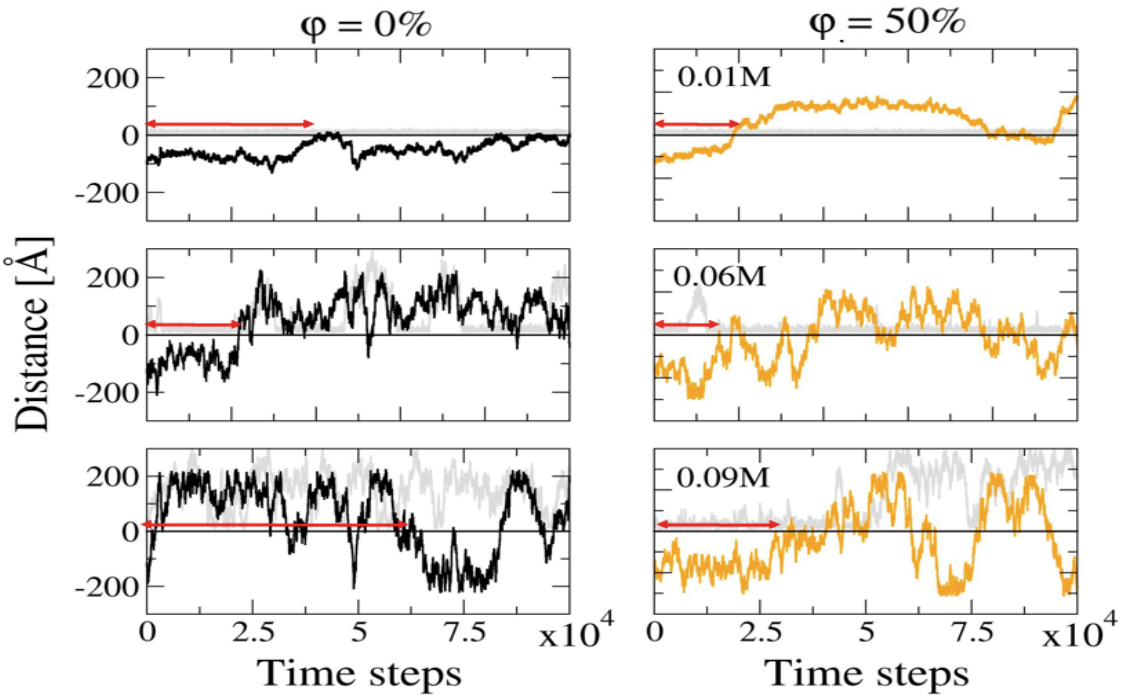

Figure A.1 – Raw trajectories data of the displacement of the protein during 10^5 time steps along the DNA axis under salt concentrations of 0.01M (upper panels), 0.06M (middle panels) and 0.09M (lower panels) at $\varphi = 0\%$ (left panels, black lines) and $\varphi = 50\%$ (right panels, orange lines). At each trajectory, protein displacement from DNA axis is shown as grey lines. At each trajectory red arrow represents the value of τ^{Target}. Corresponding to Fig. 4.3A in the main text, τ^{Target} decreases with increasing fractional volume, φ, for all three salt concentrations.

shown in Figure A.4C, diffusion of the DBP is highly hindered in the bulk and the finding time increases.

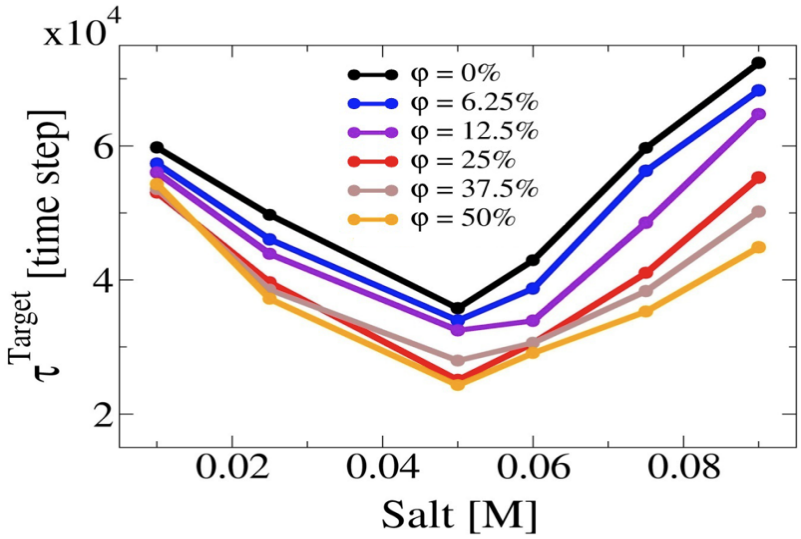

Figure A.2 – The effect of molecular crowding (modeled as φ, the fractional volume) on τ^{Target}, when placing the protein far from the DNA, at different salt concentrations. Although this results in higher values of τ^{Target}, the overall shape resembles that of Fig. 4.3A in main text, with τ^{Target} decreasing with increasing fractional volume, φ, for all salt concentrations.

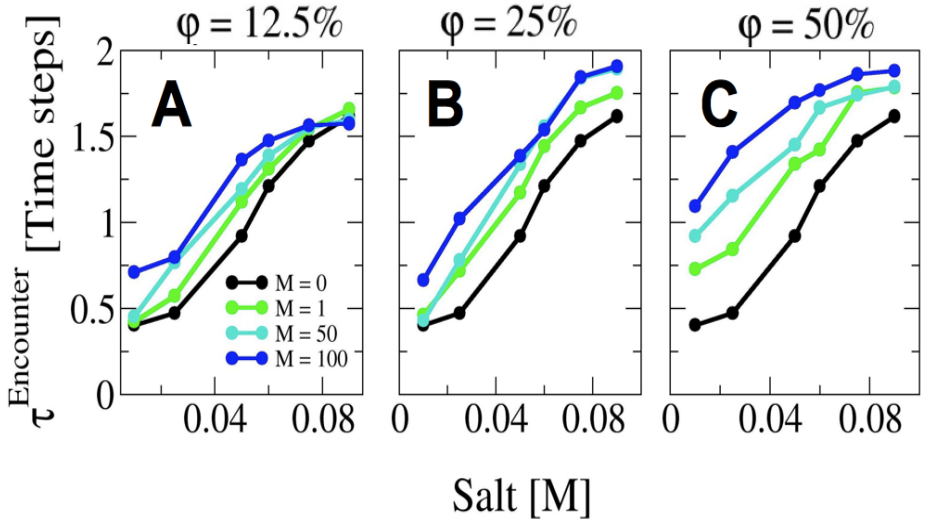

Figure A.3 – The effect of molecular crowding mass on τ^{Target} when placing the protein far from the DNA at fractional volumes of $\varphi = 12.5\%$ (left panel), $\varphi = 25\%$ (middle panel) and $\varphi = 50\%$ (right panel), as a function of salt concentrations. Corresponding to Fig. 4.4A in main text, the slow movement of crowders induces the effect of confinement, resulting in the increase of $\tau^{\text{Encounter}}$ with increasing fractional volume, φ, for all salt concentrations.

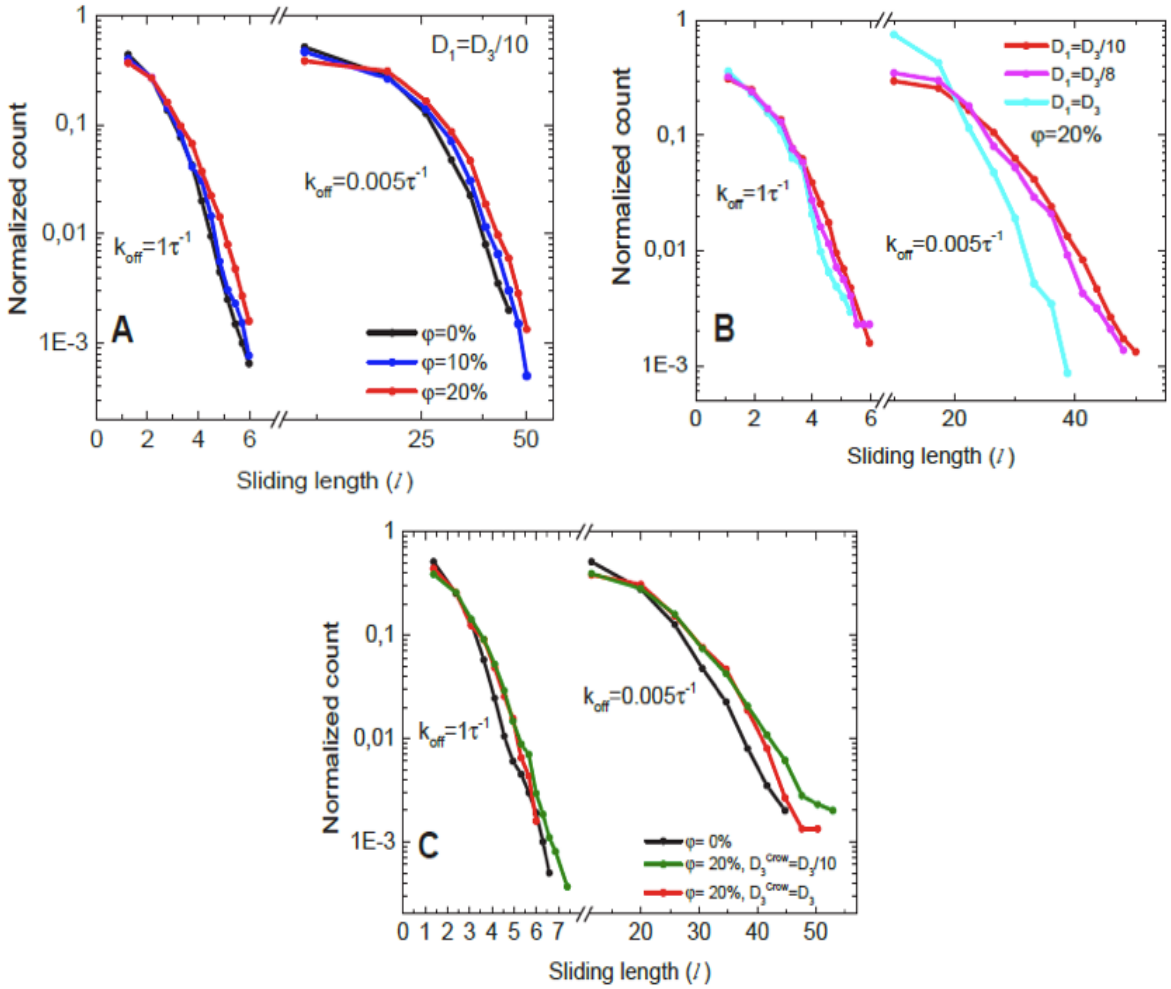

Figure A.4 – Sliding length distributions of the DBP along the DNA for various values of k_{off},D_1 and φ. A) As molecular crowding increases, DBP-DNA binding is enhanced and the sliding length distributions shift towards larger values. B) As the 1D diffusion constant increases, the distributions get narrow around a smaller value of sliding length, because scanning of lattice sites become faster. C) In addition to the enhancement of DBP-DNA binding by crowding, heavy crowders increase the effect. Sliding length distributions shift towards higher values. We note that despite the enhancement in DBP-DNA binding, the strong effects that heavy crowders have on 3D diffusion make that the average finding time drastically increases (Fig. 4.5B main text). All data shown here correspond to $\varphi = 20\%$.

Bibliography

[1] S. B. Zimmerman and S. O. Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J. Mol. Biol. 222:599-620, 1991.

[2] A. B. Fulton. How crowded is the cytoplasm? Cell. 30:345-347, 1982.

[3] B. Alberts. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell, 92(3):291-4,1998.

[4] H-X Zhou, G. Rivas and A. P. Minton. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. Annu. Rev. Biophys. 37:375-397, 2008.

[5] R. Phillips, J. Kondev and J. Theriot. Physical biology of the cell. Garland Science, 2008, chapter 14.

[6] R. J. Ellis. Macromolecular crowding: obvious but underappreciated. Trends in Biochem. Sci. 26(10):597-604, 2001.

[7] S. B. Zimmerman and A. P. Minton. Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu. Rev. Biophys. Biomol. Struct. 22:27-65. 1993.

[8] A. P. Minton. Excluded volume as a determinant of macromolecular structure and reactivity. Biopolymers 20:2093-2120, 1981.

[9] B. van den Berg, R. J. Ellis and C. M. Dobson. Effects of macromolecular crowding on protein folding and aggregation. The EMBO J. 18(24):6927-6933, 1999.

[10] M. S. Cheung, D. Klimov and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. Proc. Natl. Acad. Sci. U.S.A. 102(13):4753-4758, 2005.

[11] J. Mittal and R.B. Best. Dependence of Protein Folding Stability and Dynamics on the Density and Composition of Macromolecular Crowders. Biophys. J. 98:315-320, 2010.

[12] Y. C. Kim, R. B. Best and J. Mittal. Molecular crowding effects on protein-protein binding affinity and specificity. J. Chem. Phys. 133(20):205101, 2010.

[13] H. Matsuda, G. G. Putzel, V. Backman and I. Szleifer. Macromolecular crowding as a regulator of gene transcription. Biophys. J. 106:1801-1810, 2014.

[14] M. J. Morelli, R. J. Allen and P. R. ten Wolde. Effects of macromolecular crowding on genetic networks. Biophys. J. 101:2882-2891, 2011.

[15] C. Tan, S. Saurabh, M. P. Bruchez, R. Schwartz and P. LeDuc. Molecular crowding shapes gene expression in synthetic cellular nanosystems. Nature Nanotechnology. 8:602-608, 2013.

[16] S. R. McGuffee, A. H. Elcock. Diffusion, crowding and protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comp.l Biol. 6(3)e1000694, 2010.

[17] E. Roberts, A. Magis, J. O. Ortiz, W. Baumeister and Z. Luthey-Schulten. Noise contributions in an inducible genetic switch: A whole-cell simulation study. PLoS Comp. Biol. 7(3)e1002010, 2011.

[18] S. L. Cravens, J. D. Schonhoft1, M. M. Rowland, A. A. Rodriguez, B. G. Anderson and J. T. Stivers. Molecular crowding enhances facilitated diffusion of two human DNA glycosylases. Nucleic Acids Res. 43(8):4087-4097, 2015.

[19] D. Gnutt, M. Gao, O. Bryski, M. Heyden, and S. Ebbinghaus. Effekte des Volumenausschlusses in lebenden Zellen. Angew. Chem. 127: 2578 -2581, 2015.

[20] B. K. Derham and J. J. Harding. The effect of the presence of globular proteins and elongated polymers on enzyme activity. Biochim. Biophys. Acta. 1764:1000-1006, 2006.

[21] M. G. S. Norris and N. Malys. What is the true enzyme kinetics in the biological system? An investigation of macromolecular crowding effect upon enzyme kinetics of glucose-6-phosphatase dehydrogenase. Biochem. Biophys. Res. Commun. 405:388-392, 2011.

[22] N. Muramatsu and A. P. Minton. Tracer diffusion of globular proteins in concentrated protein solutions. Proc. Nat. Acad. Sci. USA. 85:2984-2988, 1988.

[23] J. Han and J. Herzfeld. Macromolecular diffusion in crowded solutions. Biophys. J. 65:1155-1161, 1993.

[24] J. S. Kim and A. Yethiraj. Effects of macromolecular crowding on reaction rates: a computational and theoretical study. Biophys. J. 96:1333:1340, 2009.

[25] D. Gomez and S. Klumpp, Biochemical reactions in crowded environments: Revisiting the effects of volume exclusion with simulations, Front. Phys. 3, 45 (2015).

[26] D. F. Browning and S. J. Busby. The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. 2(1):57-65, 2004.

[27] G. D. Stormo. DNA binding sites: representation and discovery. Bioinformatics. 16(1):16-23, 2000.

[28] C. A. Brackley, M. E. Cates, and D. Marenduzzo. Intracellular Facilitated Diffusion: Searchers, Crowders, and Blockers. Phys. Rev. Lett. 111:108101, 2013.

[29] S. P. Haugen, W. Ross and R. L. Gourse. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat. Rev. Microbiol. 6: 507-519, 2008.

[30] E. Nudler. RNA polymerase active center: the molecular engine of transcription. Annu. Rev. Biochem. 78:335-361, 2009.

[31] R. M. Saecker, M. T. Record and P. L. Dehaseth. Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol. 412:754-771, 2011.

[32] 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.

[33] A. Marcovitz and Y. Levy. Obstacles may facilitate and direct DNA search by proteins. Biophys. J. 104:2042-2050, 2013.

[34] R. J. Ellis and A. P. Minton. Join the crowd. Nature. 425:27-28, 2003.

[35] S. Asakura and F. Oosawa. On interaction between two bodies immersed in a solution of macromoleculars. J. Chem. Phys. 22:1255-1256, 1954

[36] M. V. Smoluchowski. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. Z. Phys. Chemie 92:129 -168, 1917.

[37] M. T. Moran-Zorzano, A. Viale, F. Munoz, N. Alonso-Casajus, G. G. Eydallin, B. Zugasti, E. Baroja-Fernandez and J. Pozueta-Romero. Escherichia coli AspP activity is enhanced by macromolecular crowding and by both glucose-1,6-bisphosphate and nucleotide-sugars. FEBS Lett. 581:1035-40, 2007.

[38] K. Snoussi and B. Halle. Protein self-association induced by macromolecular crowding: a quantitative analysis by magnetic relaxation dispersion. Biophys. J. 88:2855-66, 2005

[39] N. Kozer, Y. Y. Kuttner, G. Haran and G. Schreiber. Protein-protein association in polymer solutions. Biophys. J. 92:2139-49, 2007.

[40] N. Kozer and G. Schreiber. Effect of crowding on protein-protein association rates: fundamental differences between low and high mass crowding agents. J. Mol. Biol. 336:763-74, 2004.

[41] Y. Y. Kuttner, N. Kozer, E. Segal, G. Schreiber and G. Haran. Separating the contribution of translational and rotational diffusion to protein association. J. Am. Chem. Soc.127:15138-44, 2005.

[42] T. Diaz-Lopez, C. Davila-Fajardo, F. Blaesing, M. P. Lillo and R. Giraldo. Early events in the binding of the pPS10 replication protein RepA to the single iteron and operator DNA sequences. J. Mol. Biol. 364:909-20, 2006.

[43] M. Jiang, Z. H. Guo. Effects of macromolecular crowding on the intrinsic catalytic efficiency and structure of enterobactin-specific isochorismate synthase. J. Am. Chem. Soc. 129:730-31, 2007.

[44] F. Crick. Central dogma of molecular biology. Nature. 227:561-563, 1970.

[45] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira. Molecular cell biology. W. H. Freeman, 2004.

[46] U. Alon. An introduction to systems biology. Chapman & Hall, 2007.

[47] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon. Network motifs: simple building blocks of complex networks. Science. 298(5594):824-7, 2002.

[48] S. S. Shen-Orr, R. Milo, S. Mangan and Uri Alon. Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics. 31, 64-68, 2002.

[49] M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain. Stochastic gene expression in a single cell. Science 297, 1183-1186 (2002).

[50] M. B. Elowitz, and S. Leibler. A synthetic oscillatory network of transcriptional regulators. Nature, 403, 335 -338, 2000.

[51] M. Scott, T. Hwa and B. Ingalls. Deterministic characterization of stochastic genetic circuits. Proc. Nat. Acad. Sci. USA. 104:7402-7407, 2007.

[52] J. Hasty, F. Isaacs, M. Dolnik, D. McMillen, J.J. Collins. Designer gene networks: Towards fundamental cellular control. Chaos, 11(1): 207-220, 2001.

[53] T. S. Gardner, C. R. Cantor and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. Nature. 403(6767):339-42, 2000.

[54] A. Raj, C. S. Peskin, D. Tranchina, D. Vargas, and S. Tyagi. Stochastic mRNA synthesis in mammalian cells. PloS Biol. 309, 1707-1719 (2006

[55] B. Munsky, G. Nauert, A. van Oudenaarden. Using gene expression noise to understand gene regulation. Science. 336, 183-187, 2012.

[56] J. Paulsson. Summing up the noise in gene networks. Nature, 427:415-418, 2004.

[57] J. Paulsson. Models of Stochastic Gene Expression. Phys. Life Rev. 2:157-75, 2005.

[58] F. Isaacs, J. Hasty, C. Cantor, J.J. Collins. Prediction and measurement of an autoregulatory genetic module. Proc. Nat. Acad. Sci. USA. 100(13):7714-7719, 2003.

[59] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. U.S.A. 98:8614-8619, 2001.

[60] N. A. Cookson, S. W. Cookson, L. S. Tsimring and Jeff Hasty. Cell cycle-dependent variations in protein concentration. Nucleic Acids Research, 3(38):1-6, 2009.

[61] S. Cookson, N. Ostroff, W. Lee Pang, D. Volfson and J. Hasty. Mol. Syst. Biol. 1:2005.0024, 2005.

[62] D. Huh and J. Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat. Genet. 43(2):95-100, 2011.

[63] V. Bierbaum and S. Klumpp, Impact of the cell division cycle on gene circuits, Phys. Biol. 12, 066003 (2015).

[64] D. Gomez, R. Marathe, V. Bierbaum, and S. Klumpp, Modeling stochastic gene expression in growing cells, J. Theor. Biol. 348, 1-11 (2014).

[65] K. Zakrzewska and R. Lavery. Towards a molecular view of transcriptional control. Curr. Opin. Struct. Biol. 22:160-167, 2012.

[66] F. Wang, S. Redding, I. J. Finkelstein, J. Gorman, D. R. Reichman and E. C. Greene. The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. Nat. Struct. Mol. Biol. 20(2):174-81, 2013.

[67] L. Postow, C. D. Hardy, J. Arsuaga and N. R. Cozzarelli. Topological domain structure of the Escherichia coli chromosome. Genes & Dev. 18:1766-1779, 2004.

[68] A. Riggs, S. Bourgeois and M. Cohn. The lac repressor-operator interactions. III. Kinetic studies. J. Mol. Biol. 53:401,1970.

[69] 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 20:6929-6948, 1981.

[70] P. H. von Hippel and O. G. Berg. Facilitated target location in biological systems. J. Biol. Chem. 264:675?678, 1989.

[71] Y. M. Wang, R. H. Austin and E. C. Cox. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. Phys. Rev. Lett. 97:048302, 2006

[72] S. G. Erskine and S. E. Halford. Reactions of the EcoRV restriction endonuclease with fluorescent oligodeoxynucleotides: identical equilibrium constants for binding to specific and non-specific DNA. J. Mol. Biol. 275:759-772, 1998.

[73] S. Cranz, C. Berger, A. Baici, I. Jelesarov and H. R. Bosshard. Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site. Biochemistry, 43:718-727,2004.

[74] 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.

[75] R. K. Das and A. B. Kolomeisky. Facilitated search of proteins on DNA: correlations are important. Phys. Chem. Chem. Phys. 12:2999-2004, 2010.

[76] S. E. Halford and J. F. Marko. How do site-specific DNA-binding proteins find their targets?. Nucleic Acids Res. 32(10):3040-52, 2004.

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

[78] A. Marcovitz and Y. Levy. Sliding dynamics along DNA: a molecular perspective. Innovations in Biomolecular Modeling and Simulation, RSC, Biomolecular Sciences, 24, 10: 237-262, 2012.

[79] I. M. Sokolov, R. Metzler, K. Pant and M. C. Williams. Target search of N sliding proteins on a DNA. Biophys. J. 89: 895, 2005.

[80] M. Bauer and R. Metzler. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. Biophys. J. 102:2321, 2012.

[81] K. V. Klenin, H. Merlitz, J. Langowski and C.-X. Wu. Facilitated diffusion of DNA-binding proteins. PRL 96:018104, 2006.

[82] M. Slutsky and L. A. Mirny. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. Biophys J. 87(6):4021-35, 2004.

[83] H. Kabata, O. Kurosawa, I. Arai, M. Washizu, S. A. Margarson, R. E. Glass and N. Shimamoto. Visualization of single molecules of RNA polymerase sliding along DNA. Science. 262(5139):1561-3, 1993.

[84] N. P. Stanford, M. D. Szczelkun, J. F. Marko and S. E. Halford. One- and three-dimensional pathways for proteins to reach specific DNA sites. EMBO J. 19(23):6546-57, 2000.

[85] S. Reddinga and E. C. Greene. How do proteins locate specific targets in DNA?. Chem. Phys. Lett. 570:1-11, 2013.

[86] L. J. Friedman, J. P. Numm and J. Gelles. RNA polymerase approaches its promoter without long-range sliding along DNA. Proc. Natl. Acad. Sci. U.S.A. 110(24): 9740-9745, 2013.

[87] D. S. Cayley and M. T. Record Jr. Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate proteinDNA interactions and growth rate in osmotically stressed Escherichia coli K-12. J. Mol. Recongnition 17:488-496, 2004.

[88] G.-W. Li, O. G. Berg and J. Elf. Effects of macromolecular crowding and DNA looping on gene regulations kinetics. Nat. Phys. 5:294-297, 2009.

[89] Y. Meroz, I. Eliazar and J. Klfter. Facilitated diffusion in a crowded environment: from kinetics to stochastics. J. Phys. A: Math. Theor. 42:434012, 2009.

[90] A. Mahmutovic, O. G Berg and J. Elf. What matters for lac repressor search in vivo - Sliding, Hopping, Intersegment Transfer, Crowding on DNA or Recognition?. Nucleic Acids Res. 43(7):3454-3464, 2015.

[91] L. Liu and K. Luo. Molecular crowding effect on dynamics of DNA-binding proteins search for their targets. J. Chem. Phys. 141:225102, 2014.

[92] R. T. Dame. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol. Microbiol. 56(4):858-70, 2005.

[93] M. A. Lomholt, B. van den Broek, S. M. Kalisch, G. J. Wuite and R. Metzler. Facilitated diffusion with DNA coiling. Proc. Natl. Acad. Sci. U.S.A. 106(20):8204-8208, 2009.

[94] J. Zhang. Protein-length distributions for the three domains of life. Trends Genet. 16(3):107-109, 2000.

[95] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes. Theory of protein folding: The Energy Landscape Perspective. Annu. Rev. Phys. Chem. 48:545-600, 1997.

[96] J. D. Bryngelson, J. N. Onuchic and P. G. Wolynes. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. Proteins: Struct. Funct. Genet. 21:167-95, 1995.

[97] C. M. Dobson. Protein folding and misfolding. Nature. 18:426(6968):884-90, 2003.

[98] C. M. Dobson, A. Sali and M. Karplus. Protein folding: a perspective from theory and experiment. Angew. Chem. Int. Ed. Eng. 37, 868?893 (1998).

[99] W. Kauzmann. Some factors in the interoperation of protein denaturation. Adv. Protein Chem. 16:1-64, 1959.

[100] L. R Baldwin. Protein folding: matching speed and stability. Nature. 369:183-184, 1994.

[101] J.N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci. Toward an outline of the topography of a realistic protein- folding funnel. Proc. Natl. Acad. Sci. U.S.A. 92:3626-3630,1995.

[102] J.N. Onuchic and P. G. Wolynes. Theory of protein folding. Curr. Opin. Struct. Biol. 14(1):70-5, 2004.

[103] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins: Structure, Function, and Bioinformatics, 21(3):167-195, 1995.

[104] J. A. McCammon, B. R. Gelin and M. Karplus. Dynamics of folded proteins Nature. 267:585-590, 1977.

[105] A. Cooper. Thermodynamic fluctuations in protein molecules. Proc. natn. Acad. Sci. U.S.A. 73, 2740?2741 (1976).

[106] W. N. Lipscomb. Structure and mechanism in the enzymic activity of carboxypeptidase A and relations to chemical sequence. Acc. Chem. Res. 3(3):81-89, 1970.

[107] Y. Wang and J. A. McCammon. Introduction to Molecular Dynamics: Theory and Applications in Biomolecular Modeling, in: N.V. Dokholyan (Ed.). Computational Modeling of Biological Systems, Biological and Medical Physics, Biomedical Engineering. Springer, 2012, chapter 1.

[108] V. Hornak, R. Abel, A. Okur, B. Strockbine, A Roitberg and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 65:712-725, 2006.

[109] R. Bradley and R. Radhakrishnan. Coarse-Grained Models for Protein-Cell Membrane Interactions. Polymers. 5:890-936, 2013.

[110] J. C. Shelley and M. Y. Shelley, Computer simulation of surfactant solutions, Curr. Opin. Colloid Interface Sci. 5:101-110, 2000.

[111] P. Derreumaux. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J. Chem. Phys. 111: 2301, 1999.

[112] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. J. Chem. Theory and Comput. 4(5):819?834, 2008.

[113] J. K. Noel and J. N. Onuchic. Introduction to Molecular Dynamics: Theory and Applications in Biomolecular Modeling, in: N.V. Dokholyan (Ed.). Computational Modeling of Biological Systems, Biological and Medical Physics, Biomedical Engineering. Springer, 2012, chapter 2.

[114] C. Clementi, H. Nymeyer and J. N. Onuchic. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and "En-route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. J. Mol. Biol. 298:937-953, 2000.

[115] B. Lutz, C. Sinner, G. Heuermann, A. Verma, and A. Schug. eSBMTools 1.0: enhanced native structure-based modeling tools. Bioinformatics. 29(21):2795-6, 2013.

[116] H.-X. Zhou. Protein folding and binding in confined spaces and in crowded solutions. J. Mol. Recognit. 7: 368-375, 2004.

[117] P. McPhie, Y. S. Ni and A. P. Minton. Macromolecular crowding stabilizes the molten globule form of apomyoglobin with respect to both cold and heat unfolding. J. Mol. Biol. 361:7-10, 2006.

[118] X. Ai, Z. Zhou, Y. Bai and W. Y. Choy. 15N NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native conditions. J. Am. Chem. Soc. 128:3916-17, 2006.

[119] Y. Sasaki, D. Miyoshi and N. Sugimoto. Regulation of DNA nucleases by molecular crowding. Nucleic Acids Res. 35:4086-93, 2007.

[120] F. A. Hoekstra, E. A. Golovina and J. Buitink. Mechanisms of plant desiccation tolerance. trends in Plant Sci. 6(9):431-438, 2001.

[121] A. Thalhammer, G. Bryant, R. Sulpice, and D. Hincha. Disordered Cold Regulated15 proteins protect chloroplast membranes during freezing through binding and folding, but do not stabilize chloroplast enzymes in vivo. Plant Physiol. 166:190-201, 2014.

[122] R. J. Ellis and F. -U. Hartl. Protein folding in the cell: competing models of chaperon in function. FASEB J. 10:20-26, 1996.

[123] A. P. Minton. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. Journal of Pharmaceutical Sciences. 94(8):1668-1675, 2005.

[124] J. S. Kim and A. Yethiraj. Crowding effects on protein association: effects of interactions between crowding agents. J. Phys. Chem. B. 115:347:353, 2011.

[125] J. Schöneberg and Frank Noe. Readdy - A software for particle-based reaction-diffusion dynamics in crowded cellular environments. PLos ONE. 8(9)e74261, 2013.

[126] M. Kumar, M. S. Mommer and V. Sourjik. Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. Biophys. J. 98:552-559, 2010.

[127] S. Klumpp, M. Scott, S. Pedersen and T. Hwa. Molecular crowding limits translation and cell growth. Proc. Natl. Acad. Sci. U.S.A. 110:16754-16759, 2013.

[128] I. M. Krab, A. Parmeggiani. EF-Tu, a GTPase odyssey. Biochem. et Biophys. Acta. 1443:1-22, 1998.

[129] P. Atkins and J. de Paula. Physical chemistry. Oxford University Press, 2002.

[130] J. S. Zon, M. J. Morelli, S. Tanase-Nicola, and P. R. ten Wolde, Diffusion of transcription factors can drastically enhance the noise in gene expression. Biophys. J. 91:4350-4367, 2006

[131] S. Asakura and F. Oosawa. Interaction between particles suspended in solutions of macromolecules. J. Polymer Sci. 33:183-192, 1958.

[132] L. A. Benton, A. E. Smith, G. B. Young and G. J. Pielak. Unexpected effects of macromolecular crowding on protein stability. Biochemistry. 51:9773-9775, 2012.

[133] L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. Biochem. Z. 49:333-369, 1913.

[134] K. Totani, Y. Ihara, I. Matsuo and Y. Ito. Effects of macromolecular crowding on glycoprotein processing enzymes. J. Am. Chem. Soc. 130:2101-2107, 2008.

[135] M. T. Record, W. S. Reznikoff, M. A. Craig, K. L. McQuade, and P. J. Schlax. Escherichia coli RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella*, pages 792–820. ASM Press, Washington DC, 2nd edition, 1996.

[136] Patrick P Dennis, Mans Ehrenberg, and Hans Bremer. Control of rrna synthesis in escherichia coli: a systems biology approach. *Microbiol Mol Biol Rev*, 68(4):639–668, 2004.

[137] M. Tabaka, T. Kalwarczyk and R. Holyst. Quantitative influence of macromolecular crowding on gene regulation kinetics. Nucl. Acids Res. 42: 727-738, 2013

[138] G. K. Ackers, A. D. Johnson and M. A. Shea. Quantitative model for gene regulation by λ phage repressor. Proc. Natl. Acad. Sci. U.S.A. 79:1129-1133, 1982.

[139] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev and R. Phillips. Transcriptional regulation by the numbers: models. Curr. Opin. in Gen. & Dev. 15:116-124, 2005.

[140] Y. Kao-Huang, A. Revzin, A. P. Butler, P. O'Conner, D. W. Noble, and P. H. von Hippel. Nonspecific dna binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proc Natl Acad Sci U S A*, 74:4228–4232, 1977.

[141] D. Berg and M. Chamberlin. Physical Studies on Ribonucleic Acid Polymerase from Escherichia coli B. Biochem. 9(26):5055-5064, 1970.

[142] S. Bakshi, R.M. Dalrymple, W. Li, H. Choi, and J.C. Weisshaar. Partitioning of RNA Polymerase Activity in Live Escherichia coli from Analysis of Single-Molecule Diffusive Trajectories. Biophys. J. 105, 2676-2686, 2013.

[143] B. P. Bratton, R. A. Mooney and J. C. Weisshaar. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. J. of bacteriol. 193(19):5138-5146, 2011.

[144] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. In F C Neidhardt, editor, *Escherichia coli and Salmonella*, pages 1553–1569. ASM Press, Washington DC, 2nd edition, 1996.

[145] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang and T. Hwa. Interdependence of cell growth and gene expression: origins and consequences. Science. 330:1099-1102, 2010.

[146] S. Klumpp and T. Hwa. Bacterial growth: global effects on gene expression, growth feedback and proteome partition. Curr. Opin. in Biotech. 28:96-102, 2014.

[147] H. Dong, L. Nilsson and C. G. Kurland. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J. Mol. Biol. 260:649-663, 1996.

[148] G. Cannarozzi, N N. Schraudolph, M Faty, P von Rohr, M. T. Friberg, A. C. Roth, P. Gonnet, G. Gonnet, and Y. Barral. A Role for Codon Order in Translation Dynamics. Cell 141, 355-367, 2010

[149] U. Gerland, J. D. Moroz, and T. Hwa. Physical constraints and functional characteristics of Transcription Factor-DNA interaction. Proc. Natl. Acad. Sci. USA. 99(19):12015-20, 2002.

[150] A. B. Kolomeisky. Physics of protein-DNA interactions:mechanisms of facilitated target search. Phys. Chem. Chem. Phys. 13:2088-2095, 2013.

[151] 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. 20:6961-6977, 1981.

[152] 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. USA. 102:15883-15888, 2005.

[153] C. L. Peterson and M. A. Laniel. Histones and histone modifications. Curr. Biol. 14(14):R546-51, 2004.

[154] S. C. Dillon and C. J. Dorman. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nature Reviews Microbiology. 8:185-195, 2010.

[155] O. Givaty and Y. Levy. Protein Sliding Along DNA: Dynamics and Structural Characterization. J Mol Biol. 385:1087-1097, 2009.

[156] D. M. Gowers and S. E. Halford. Protein motion from nonspecific to specific DNA by three-dimensional routes aided by supercoiling. EMBO J. 22:1410-1418, 2003.

[157] S. E. Halford. An end to 40 years of mistakes in DNA-protein association kinetics? Biochem. Soc. T 37:343-348, 2009.

[158] T. Terakawa, H. Kenzaki and S. Takada. p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains. J. Am. Chem. Soc. 134(35)14555-14562, 2012.

[159] T. Ando and J. Skolnick. Sliding of Proteins Non-specifically Bound to DNA: Brownian Dynamics Studies with Coarse-Grained Protein and DNA Models. Plos. Comput. Biol. 10(12), 2014.

[160] I. Echeverria and G. A. Papoian. DNA Exit Ramps Are Revealed in the Binding Landscapes Obtained from Simulations in Helical Coordinates. Plos. Comput. Biol. 11(2), 2015.

[161] 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. USA. 110(49):19796-19801, 2013.

[162] N. R. Zabet B. Adryan. Computational models for large-scale simulations of facilitated diffusion. Mol .Biosyst. 8(11):2815-2827, 2012.

[163] G. M. Clore. Exploring translocation of proteins on DNA by NMR. J. Biomo.l Nmr. 51:209-219, 2011.

[164] J. Iwahara, M. Zweckstetter and G. M. Clore NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. Proc. Natl. Acad. Sci. USA. 103(41):15062-15067, 2006.

[165] J. D. Schonhoft J. T. Stivers. Timing facilitated site transfer of an enzyme on DNA. Nat. Chem. Biol. 8(2):205-210, 2012.

[166] R. H. Porecha and J. T. Stivers. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. Proc. Natl. Acad. Sci. USA. 105(31):10791-10796, 2008.

[167] A. J. Pollak, A. T. Chin and N. O. Reich. Distinct Facilitated Diffusion Mechanisms by E. coli Type II Restriction Endonucleases. Biochemistry. 53(45):7028-7037, 2014.

[168] A. Tafvizi, F. Huang, A. R. Fersht, L. A. Mirny and A. M. van Oijen. A single-molecule characterization of p53 search on DNA. Proc. Natl. Acad. Sci. USA. 108(2):563-568, 2011.

[169] L. Mirny and M. Slutsky. How a protein searches for its site ob DNA: the mechanism of facilitated diffusion. J. of Physics A-Mathematical and Theoretical. 42:434013, 2009.

[170] M. Bauer and R. Metzler. In Vivo Facilitated Diffusion Model. PloS one 8(1), 2013.

[171] A. H Elcock. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. Curr. Opin. Struc. Biol. 20(2):196-206, 2010.

[172] B. Akabayov, S. R. Akabayov, S. J. Lee, G. Wagner and C. C. Richardson. Impact of macromolecular crowding on DNA replication. Nat .Commun. 4:1615, 2013.

[173] A. Bhattacharya, Y. C. Kim and J. Mittal. Protein-protein interactions in a crowded environment. Biophys. Rev. 5(2):99-108, 2013.

[174] A. Veksler and A. B. Kolomeisky. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is It Real or Not? J. of Phys. Chem. B. 117(42):12695-12701, 2013.

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

[176] M. Sheinman, O. Benichou, Y. Kafri and R. Voituriez. Classes of fast and specific search mechanisms for proteins on DNA. Rep. Prog. Phys. 75(2):026601, 2012.

[177] 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.

[178] A. Marcovitz and Y. Levy. Weak frustration regulates sliding and binding kinetics on rugged protein-DNA landscapes. J. Phys. Chem. B. 117(42):13005-13014, 2013.

[179] A. Bhattacherjee and Y. Levy. Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. Nucl. Acids Res. 42(20):12404-12414, 2014.

[180] N. Khazanov, A. Marcovitz and Y. Levy. Asymmetric DNA-search dynamics by symmetric dimeric proteins. Biochem. 52(32):5335-5344, 2013.

[181] N. Khazanov and Y. Levy. Sliding of p53 along DNA can be modulated by its oligomeric state and by cross-talks between its constituent domains. J Mol Biol 408(2):335-355, 2011.

[182] D. Vuzman, A. Azia and Y. Levy. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. J. Mol. Biol. 396(3):674-684, 2010.

[183] D. Vuzman, M. Polonsky and Y. Levy. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. Biophys. J. 99(4):1202-1211, 2010.

[184] L. Zandarashvili, D. Vuzman, A. Esadze, Y. Takayama, D. Sahu, Y. Levy and J. Iwahara. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. Proc. Natl. Acad. Sci. USA. 109(26):E1724-1732, 2012.

[185] L. Zandarashvili, A. Esadze, D. Vuzman, C. A. Kemme, Y. Levy J. Iwahara. Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. Proc. Natl. Acad. Sci. USA. 112(37):E5142-E5149, 2015.

[186] A. Mondal and A. Bhattacherjee. Searching target sites on DNA by proteins: Role of DNA dynamics under confinement. Nucl. Acids Res. 43(19):9176-9186, 2015.

[187] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu and J. N. Onuchic. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. Nucl. Acids Res. 38(Web Server issue):W657-661, 2010.

[188] E. J. Sambriski, D. C. Schwartz and J. J. De Pablo. A mesoscale model of DNA and its renaturation. Biophys. J. 96:1675-1690, 2009.

[189] T. Schlick. Molecular modeling and simulation - An interdisciplinary guide. pp 298-301, 2002.

[190] A. Mor, G. Ziv and Y. Levy. Simulations of proteins with inhomogeneous degrees of freedom: The effect of thermostats. J. of Comp. Chem. 29(12):1992-1998, 2008.

[191] Y. Ueda, H. Taketomi and N. Go. Studies on protein folding, unfolding, and fluctuations by computer Simulation. II. a three-dimensional lattice model of lysozyme. Biopolymers. 17:1531-1548, 1978.

[192] M. Levitt and A. Warshel. Computer simulation of protein folding. Nature. 253:694-698, 1975.

[193] A. D Miranker and C. M Dobson. Collapse and cooperativity in protein folding.Curr. Opin. Struct. Biol. 6:31-42, 1996.

[194] H. Nymeyer, N. D. Socci and J. N. Onuchic. Land- scape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. Proc. Natl Acad. Sci. USA. 97:634-639, 2000.

[195] J. N. Onuchic, H. Nymeyer, A. E. Garcia, J. Chahine and N. D. Socci. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. Advan. Protein Chem. 53:87–52, 2000.

[196] P. C Whitford, K. Y. Sanbonmatsu and J. N Onuchic. Biomolecular Dynamics: Order-disorder transitions and energy landscapes. Rep. Prog. Phys. 75:076601, 2012.

[197] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu and J. N. Onuchic. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. Proteins. 75: 430-441, 2009.

[198] B. Lutz, M. Faber, A. Verma, S. Klumpp, and A. Schug, Differences between co-transcriptional and free riboswitch folding, Nucl. Acids Res. 42, 2687-2696, 2014.

[199] M. Faber and S. Klumpp, Kinetic Monte Carlo approach to RNA folding dynamics using structure-based models, Phys. Rev. E 88, 052701, 2013.

[200] Y. Levy, P. G. Wolynes and J. N. Onuchic. Protein topology determines binding mechanism. Proc. Nat. Acad. Sci. USA. 101(2):511-516, 2004.

[201] M. A. Arbing, S. K. Handelman, A. P. Kuzin, G. Verdon, C. Wang, M. Su, F. P. Rothenbacher, M. Abashidze, M. Liu, J. M. Hurley, R. Xiao, T. Acton, M. Inouye, G. T. Montelione, N. A. Woychik, and J. F. Hunt. Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. Structure. 18(8):996-1010, 2010.

[202] L. Stagg, S.-Q. Zhang, M. S. Cheung and P. Wittung-Stafshede. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. Proc. Natl. Acad. Sci. USA. 104:18976-81, 2007.

[203] A. Dhara, A. Samiotakisb, S. Ebbinghausa, L. Nienhausa, D. Homouzb, M. Gruebelea, and M. S. Cheung. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. Proc. Natl. Acad. Sci. USA. 107(41):17586-17591, 2010.

[204] S. Qin and H.-X. Zhou. Atomistic Modeling of Macromolecular Crowding Predicts Modest Increases in Protein Folding and Binding Stability. Biophys J. 97(1):12-9, 2009.

[205] H.-X. Zhou. Effect of mixed macromolecular crowding agents on protein folding. Proteins. 72(4):1109-13, 2008.

[206] D. Marenduzzo, K. Finan and P. R. Cook. The depletion attraction: an underappreciated force driving cellular organization. J Cell Biol. 175(5):681-6, 2006.

[207] M. Jiao, H.-T. Li, J. Chen, A. P. Minton and Y. Liang. Attractive protein-polymer interactions markedly alter the effect of Macromolecular crowding on protein association equilibria. Biophys. J. 990(3):914-23, 2010.

[208] J. Rosen, Y. C. Kim and J. Mittal. Modest protein-crowder attractive interactions can counteract enhancement of protein association by intermolecular excluded volume interactions. J. Phys. Chem. B. 1150(11):2683-9, 2011.

[209] Y. Phillip, E. Sherman, G. Haran and G. Schreiber. Common crowding agents have only a small effect on protein-protein interactions. Biophys. J. 970(3):875-85, 2009.

[210] Y. Phillip, V. Kiss and G. Schreiber. Protein-binding dynamics imaged in a living cell. Proc. Natl. Acad. Sci. USA. 1090(5):1461-6, 2012.

[211] M. Gao, D. Gnutt, A. Orban, B. Appel, F. Righetti, R. Winter, F. Narberhaus, S. Müller and S. Ebbinghaus. RNA hairpin folding in the crowded cell. Angew. Chem. Int. Ed., 55(9):3224-8, 2016.

Aknowledgments

I would like to thank my supervisor, Stefan, for his constant support and guidance over the last few years as this work was put together. I would also like to thank my colleagues and friends from my work group and the Max Planck Institute for making my time there truly special and fun. In particular, I would like to thank Leo, Salome, Davide, Marquitos, Vasil, Rafa, Yun, Mehmet, Anja, Marco, Bahareh, Rene, Marko, Susann, Michael, Roland, Rahul, Vroni, Corina, Anjan, Mamata, Pintu, Sumi, Tom, Bastian, Jaime, Jörg, Sophia, JJ and Angelo.

I'd like to thank all the people who contributed and shared their knowledge with me or collaborated with me during this work. In particular, I'm very grateful to Koby and Dana for giving me the opportunity to collaborate with them and for making my time in Israel such a beautiful experience. I'm also very grateful to Carlos for sharing his knowledge and expertise in simulations. I thank Dirk, Anne and Anja for our enjoyable discussions and the opportunity to collaborate. Special thanks to Sonja who made this contact possible.

Muchas gracias a mis amigos y personas que han compartido conmigo experiencias maravillosas en diferentes lugares. Gracias por su apoyo incondicional y por haber estado conmigo durante estos últimos años. Igualmente, me gustaría agradecer a mi familia por todo el apoyo y amor que me han brindado desde mi partida de Colombia. Para todos ha sido un proceso difícil y de mucho aprendizaje, pero el amor que nos une y por la familia que somos hemos podido salir adelante y seguir unidos. Muchas gracias a mis papas, a mis hermanos y a pimpón por todo.

Finally, I would like to thank a person that has been involved in all this work. Although her name is not written anywhere in the thesis, without her patience and love, many of the things presented here would have been different. I thank you for all the beautiful experiences and moments we have shared during the last years. Thank you for your smile, your love, and for teaching me so many things and for making me a better person. Thanks for everything, Lyz.