Molecular Metamorphosis in Transcriptional Regulation

MALIN LÜKING
Dissertation presented at Uppsala University to be publicly examined in B21, BMC, Husargatan 3, Uppsala, Wednesday, 11 October 2023 at 09:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Victor Muñoz (UC Merced, Protein Biophysics and Bioengineering).

Abstract

The foundation of all life is the interaction of molecules. Molecular interaction occurs in the tightly packed cytoplasm. In this crowded environment, the molecules need to be able to establish stable interactions with specific interaction partners, and with specific ones only. Selectivity is essential for transcriptional regulation where DNA-binding proteins, more specifically transcription factors, need to reach their respective target DNA, such as an operator sequence, quickly. The transcription factor LacI for example reaches its DNA operator within minutes and establishes a tight interaction to it that will block transcriptions of the genes it regulates. With the other millions of sequences in the bacterial chromosome, it interacts only fleetingly. This is possible because a small part of the protein is disordered during the transient interaction with non-operator DNA. Only with specific DNA will the folding of the disordered region into a rigid helix be favored. When the disordered region folds into a helix, the transcription factor changes its structure and, with it, its function from a weak DNA binder to a strong one. Thinking about molecular structures as dynamic entities that react to their environment and understanding the origins of structural metamorphosis allow us to predict changes in the protein sequence that will affect their behavior in the cellular context. This is demonstrated in this thesis where I study the structural dynamics of LacI with molecular dynamic simulations and use the insights from these simulations to design protein variants with changed binding stability and selectivity.

Keywords: Protein-DNA interactions, Facilitated diffusion, Binding Kinetics, Molecular Simulations, Fluorescence Microscopy

Malin Lüking, Department of Cell and Molecular Biology, Molecular Systems Biology, Box 596, Uppsala University, SE-751 24 Uppsala, Sweden.

© Malin Lüking 2023

ISSN 1651-6214
ISBN 978-91-513-1882-0
URN urn:nbn:se:uu:diva-509935 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-509935)
Für Mama
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the publishers.
# Contents

1 Introduction ......................................................................................... 9

2 Motivation and Aims of this Thesis .................................................. 12

3 Molecular Driving Forces and How to Visualize Them ....................... 13
   3.1 Molecules and Their Interactions ................................................. 13
   3.2 How to Model and Modulate Binding Kinetics? ......................... 15
      3.2.1 Protein Structure, Disorder and Complex Formation .......... 15
      3.2.2 A Short Introduction to DNA Structure ............................. 18

4 Protein-DNA Interactions in Context ................................................ 22
   4.1 Facilitated Diffusion ................................................................. 22
      4.1.1 A Note on Specificity and Non-Specific Binding ............... 27
   4.2 The Two-State Model ............................................................... 28
   4.3 Genes and Genetic Regulation - The lac Operon ...................... 28
   4.4 Structures of LacI - The Role of Flexibility ............................... 30

5 Methods ............................................................................................. 33
   5.1 Molecular Mechanics ............................................................... 33
      5.1.1 All-Atom Molecular Models ............................................. 34
      5.1.2 Coarse-Grained Molecular Models ................................... 37
   5.2 Molecular Dynamics ............................................................... 39
      5.2.1 Modelling Diffusive Processes ....................................... 40
   5.3 Enhanced Sampling and Metadynamics .................................... 41
   5.4 Combining Simulation and Experiments through Rational Design .. 44

6 Results ............................................................................................... 46
   6.1 Correlated Motions as Stabilizers of Specific Interactions
      (Paper I) ..................................................................................... 46
   6.2 Protein Sliding and Hopping on DNA (Paper II) ......................... 49
   6.3 Enhanced Sampling of Specific Recognition (Paper III) ............. 51
   6.4 Conformational Switching and Recognition in the lac operon
      Paper IV ..................................................................................... 53

7 Conclusions ....................................................................................... 54

8 Deutsche Zusammenfassung .............................................................. 55

9 Svensk Sammanfattning .................................................................... 57
1. Introduction

"I will keep my one atom of faith,
    said my soul.
From *The Story of Mary MacLane* by Mary MacLane

Two textbook principles guide our understanding of molecular interactions: Firstly, that the structures of molecules inform us about their functions. Secondly, that the number of contacts formed between two molecules determine the stability (the lifetime) of their complexes [3, 108]. In my thesis, I shift the focus from the molecular complex to the process of complex formation to study molecular recognition in biology. Following a long-standing idea, my results show that flexibility and structural re-arrangements (*conformational changes*\(^1\)) within the interacting bio-molecules determine their ability to form strong complexes with high specificity.

From the start, I want to be clear on my usage of terms related to binding and recognition: I define molecular recognition as the process that precedes the formation of a long-lived complex between two molecules. I will call such a long-lived complex a *specific complex*. In contrast, I will talk about *non-specific* interactions. Those interactions are short-lived (transient) and can essentially form between any two molecules. I use the term *binding* for the formation of non-specific and specific complexes but usually, I refer to the formation of a specific complex as *specific binding*. Finally, *specificity* is the difference in stability between specific and non-specific complexes.

Why is specificity so important in biology? Tens of millions of molecules are packed tightly together in a cell and perform chemical processes that ensure survival. Any too stable or too weak interaction has the potential to kill the organism. Understanding the stability, kinetics, and specificity of complex formation is therefore of great interest in medicine, pharmacology, and biotechnology [43, 38].

Proteins are the molecules that perform almost all tasks in a cell, whereas deoxyribonucleic acid (DNA) encodes the building plans (genes) for those proteins. The regulation of information flow from genes to proteins includes transcriptional and translational regulation. Transcriptional regulation is performed by so-called *transcription factors (TFs)* [161]. Due to this mechanism, cells can thrive in different environments [111]. TFs are one class of

---

\(^1\)no bonds are broken or formed, in contrast to *configurational changes* that involve bonds breaking and forming
**DNA-binding proteins (DBPs).** DBPs perform a number of tasks other than transcriptional regulation. Examples include the single-stranded DNA binding proteins, that protect DNA while it is read or copied, and proteins which detect and repair DNA damage, e.g. LexA and RecA [112]. These examples demonstrate that the specific and fast complex formation between proteins and DNA is vital. But how does it work?

I study this question on one particular DBP, namely LacI of *Escherichia coli*. This TF, and the genes it regulates, the *lac* operon, were first described in the famous work of Jacob and Monod [76]². LacI specifically binds three short stretches of DNA with a particular sequence, its operators. The specific protein-DNA complex formed between LacI and its operators lasts for several minutes [65]. Transient interactions with other DNA sequences, non-specific sites for LacI, only last fractions of a second. This is crucial as non-specific sites are in large excess (10⁶ times). With only a few LacI molecules (about 10 [108]), off-target site trapping at non-specific sites is fatal.

The theoretical determination of the rate constants for the LacI-operator complex formation has a long history in science. When first measured in the 1970ies, fast specific association rate constants between LacI and its operator sequences caused many researchers to wonder how speedy search and, simultaneous strong binding could be attained by one molecule [132, 154]. Today, fast search is explained by *facilitated diffusion* [182, 173] (see Chapter 4).

Still, central questions remain about the process of recognition at the specific site. Their answers are tricky to find because only the specific complex is well characterized, whereas the transition from non-specific to specific complexes is not [142]. Studying this transition requires knowledge about alternative protein conformations during non-specific interaction [154], the coupling of folding and specific binding [52], and interactions in the complex that precedes the formation of the specific complex [142].

In this place, I define my usage of the term *encounter complex* as the first complex formed by LacI and a particular DNA stretch. I regard the encounter complex as a non-specific complex in which recognition and a subsequent transition into the specific complex by conformational changes can occur. I will explain this process in Chapter 3.

Structural biology offers techniques that are well suited for the study of stable complexes including X-ray crystallography [148], cryo-electron microscopy (cryo-EM) [7], and nuclear magnetic resonance (NMR) spectroscopy [73]. For the study of non-specific complexes, which can undergo larger conformational changes, experimental structural biology reaches its limits. Theoretical molecular modelling by molecular dynamics simulations (MD) can evolve a molecular structure in time. As demonstrated in figure 1.1, the time scales for processes that can be simulated are limited depending on the level of molecular description. Conformational changes are, at this moment, at

---

²Nobel Prize for Physiology or Medicine 1965
Figure 1.1. Times and Sizes related to molecular interactions in the cell and methods to study them theoretically and experimentally. Quantum Mechanics (QM) describes the electronic structure of atoms and small molecules such as water or single amino acids. Molecules as large as proteins and DNA are described with mechanistic model (for details see Chapter 5). Processes such as secondary structure formation and conformational changes can be simulated in All-Atom Molecular Dynamics (MD). Such fast processes can be resolved by fluorescence correlation spectroscopy (FCS) [187]. Processes on longer time-scales, such as diffusion, require a lower level description of the molecules involved that is achieved by coarse-grained MD. Single-particle tracking (SPT) and Förster-resonance energy transfer (FRET) [87] can be used to study diffusion experimentally.

the upper limit of all-atom descriptions. Coarse-grained molecular models instead reach longer time scales such that they allow the simulation of molecular diffusion. Just as theoretical models, experimental techniques, such as fluorescent microscopy (and others shown in Fig. 1.1), cover different time and size scales. The combinations of different techniques spanning those regimes are necessary to understand the full picture of protein-DNA recognition from the molecular to the cellular level. I experienced this fact in my work on LacI, which greatly profited from joining an experimental group\(^3\).
2. Motivation and Aims of this Thesis

Before I dive into the details of molecular interaction and genetic regulation, I want to name the aim of this project and the main challenges that were identified and, to some extent, addressed in the four papers.

The initial goal of my PhD was to identify how LacI can recognize its specific binding partners, the operators, amongst the large access of non-specific DNA. Therefore, we wanted to obtain the kinetics of the specific binding between LacI and different DNA molecules by molecular dynamics simulations.

Paper I focuses on MD simulations of the specific complex, the free DNA, and the free protein based on available X-ray structures. The results gave us insights on the complex stabilization by the specific sequence as well as the conformational spaces sampled by the two free molecules. This study was informative but the recognition between protein and DNA remained unclear. Simulations of the transitions between non-specific and specific binding are required to understand how LacI recognizes specific DNA sites. But, such simulations face the challenge posed by the lack of structural data on the encounter complex.

Approaches to the study of non-specific interaction and encounter between LacI and DNA therefore comprise a large part of this work. I study non-specific interaction and recognition in coarse-grained MD simulation (Paper II) and with enhanced sampling in all-atom MD (Paper III). We finally translate the resulting hypothesis into a study that combines all-atom MD with mutational studies, in vitro and in vivo, by fluorescent microscopy (Paper IV). The study demonstrates that LacI’s binding kinetics and specificity can be substantially changed by targeting the region of conformational switching with point mutations. By combining a simulation-based hypothesis with experiments, we open up a way of studying, and eventually predicting, bio-molecular recognition on the cellular level. And come a few steps closer to the initial goal of my PhD work.
3. Molecular Driving Forces and How to Visualize Them

A fuzzy fellow, without feet -  
Yet doth exceeding run!  
Of velvet, is his Countenance -  
And his Complexion, dun!  

From A fuzzy fellow, without feet, by Emily Dickinson

In this chapter, I will introduce the forces that shape molecules and molecular complexes. Most of this chapter focuses on proteins and protein folding. The reason for this is twofold. Firstly, the molecular origins of protein-folding and protein-DNA recognition are the same, where theories on protein folding are already well-established. Secondly, there is strong evidence that protein-DNA recognition is coupled to protein folding. Energy landscapes can be used to conceptualize coupled folding and recognition. I will develop a hypothetical landscape for the recognition of specific sequences by LacI that I use to put the results of this thesis into context. Finally, I discuss the structure and potential markers of specificity in DNA.

3.1 Molecules and Their Interactions

Molecules emerge when atoms form covalent bonds. Those bonds have energies ranging from about 35 to just above 260 kcal·mol⁻¹ that is released when breaking the bond [4]. Most biomolecules are polymers of smaller molecules: Polypeptides, such as proteins, are chained amino acids connected by peptide bonds. Polypeptides can have many sizes, from a few amino acids to thousands. Nucleotides on the other hand polymerize to DNA and RNA by forming phosphodiester bonds [3]. The E. coli DNA, the circular chromosome, consists of about 10 million deoxyribonucleotides [48], most RNAs of just a few tens of ribonucleotides [147].

Polypeptides and nucleic acid chains assemble into more complex, 3-dimensional structures. These structures are usually not stabilized by covalent, but by non-covalent bonds [3]. Just as covalent bonds, non-covalent bonds originate in the electronic structure of the atoms and more precisely in their electronegativity and polarizability. Of the common elements in biology, oxygen has the highest electronegativity, followed by nitrogen and carbon [4]. The sharing, loss, or gain of valence electrons due to an element’s electronegativity
leaves molecules with partial and formal charges, polar or charged. Polarity and charge are the origin of most non-covalent bonds.

The strongest non-covalent interactions with energies of up to 3 kcal \(\cdot\) mol\(^{-1}\) in water [3] are \textit{electrostatic interactions} and \textit{hydrogen bonds}. \textit{Electrostatic interactions} between formally charged atoms are long-range interactions. Such interactions are known to be important for the kinetics of specific complex formation between proteins and proteins [171, 146] and proteins and DNA [89, 31]. In apolar environments, such as the interior of a protein, the breaking of electrostatic interactions can require much larger energies.

\textit{Hydrogen bonds} form between partially negatively charged atoms with high electronegativity (hydrogen bond acceptor) and hydrogen with am electron deficits (hydrogen bond donors). Hydrogen bonds are short-range interactions and energies from 2-10 kcal \(\cdot\) mol\(^{-1}\) [42] are needed to break them. They are usually weak in aqueous solution [3]. Charges and hydrogen bond donors and acceptors are generally fewer in apolar environments such as protein interiors and biomolecular interfaces. If the patterns of interaction partners match at interfaces between molecules though they can be markers for recognition and the origin of specificity in bio-molecular interactions [42, 92].

The largest fraction of biomolecules are hydrocarbons and they are neither charged nor polar. Such non-polar compounds interact via the so-called \textit{van der Waals (vdW) interaction} or \textit{dispersion energy}. For a single interaction, the vdW energy is lower than the thermal fluctuation (about 0.6 kcal \(\cdot\) mol\(^{-1}\)). Therefore, many such interactions have to come together at large interaction surfaces to supply sufficient energy for complex formation [3].

A special type of interaction is formed by circular hydrocarbons with delocalized electrons, that can also be found in DNA (see Fig. 3.2). This so-called \(\pi\)-\(\pi\)-stacking stabilizes the DNA structure and can also play a role in protein-DNA interactions.

Another important effect for biomolecular organization is the \textit{hydrophobic effect}, which is often named as the driving force for protein folding [42]. To explain this effect, I first introduce the concept of \textit{free energy}. So far, I have discussed energies of bonds, which are enthalpic contributions to the total energy of a system. In fact, the total or \textit{free energy} change, \(\Delta G\) (we will use Gibb’s free energy for constant pressure \((P)\) and temperature \((T)\)), of any binding event (or chemical reaction) has two parts: the change in \textit{enthalpy} \((\Delta H)\) and in \textit{entropy} \((\Delta S)\), related to the disorder or multiplicity of the system. For a reaction to occur spontaneously, \(\Delta G\) needs to be negative, which is achieved by freed (negative) enthalpy and/or positive entropy, as equation 3.1 demonstrates. According to the 2nd law of thermodynamics, the overall entropy, which is related to the number of accessible micro-states \(W\) (see Eq. 3.1, where \(R\) is the gas constant), always grows [43].

Based on this, I introduce one model that explains the hydrophobic effect, the iceberg-model [156]. Because water is polar, it forms transient hydrogen bond networks, which are enthalpically and entropically favoured. Whereas a
polar solute interacts with water via hydrogen bonds, apolar or hydrophobic molecules do not. This results in the formation of ordered water cages around the apolar molecules and lowers the entropy of the system. The aggregation of apolar hydrocarbons in water is entropically favoured because the hydrocarbon chains are already ordered. Their interaction with each other has no or a very small cost in the form of a positive entropy change [4] compared to their interaction with water. This results in a collapse of linear hydrocarbons in water caused by the hydrophobic effect. The hydrophobic collapse is the first step of protein folding, and is important for complex formation between proteins, between proteins and ligands, such as DNA, or the formation of membranes [162].

\[ \Delta G = \Delta H - T \Delta S \]  
(3.1)

\[ S = R \ln(W) \]  
(3.2)

3.2 How to Model and Modulate Binding Kinetics?

3.2.1 Protein Structure, Disorder and Complex Formation

Polypeptides fold into distinct 3-dimensional shapes. The different chemical properties of the residue backbone, polar, hydrophobic, or charged, result in a native protein conformation that is based on the amino acid sequence [42, 8]. Two different ways of forming hydrogen bonds between backbone atoms lead to the formation of two types of so-called secondary structures, namely \( \alpha \)-helices and \( \beta \)-sheets [3], which further assemble into protein domains with conserved folds (tertiary structure). Proteins often have multiple domains that can be built from one chain or several chains (quaternary structure). Outside of the structured regions, parts of the polypeptide chain remain flexible, allowing for conformational changes that are crucial for biomolecular signalling [119]. These regions are called loops or disordered regions.

Some loops can fold into secondary structures upon binding [185, 134]. Folding upon binding is now known to be a central part of protein binding to small molecules [75, 67], other proteins, [74] and DNA [120, 89]. Structural transitions induced by interaction, such as folding upon binding, are known as molecular switches [58]. I am especially interested in the folding switch in LacI [157, 168] and how it is triggered by interactions with certain DNA sequences.

It has been shown that binding kinetics can be altered by changing a protein’s preference to occupy certain conformational states [180, 161], also in protein-DNA recognition [189]. It is even possible to speed up association without affecting the stability of the specific complex [180]. This shows that
understanding the structural dynamics (the occupancy of different conformational states) of LacI and the factors that influence these dynamics can be used to engineer the TF’s specific binding kinetics in the cellular environment.

In this chapter, I explain my view on the formation of non-specific and specific LacI-DNA complexes. The details of non-specific interaction and its role in the observed search kinetics will be discussed in the next chapter. I want to note that both topics are tightly connected.

Protein folding [24, 184, 135] as well as recognition processes between proteins and DNA [150, 89] have been visualized on energy landscapes. The folding funnel, that has become famous for reasoning about Levinthal’s paradox for protein folding [51], is shown in figure 3.1 A. Levinthal’s paradox essentially describes the conundrum between theoretical estimations for the speed of protein folding and how fast it actually occurs. It has parallels to the problem of protein-DNA recognition, as I will explain in Chapter 4. Wolynes and co-authors visualize the different levels of downhill folding towards a native structure [184] using the gradual formation of native contacts (Q) to describe the molecular conformation. The number of native contacts is a so-called collective variable (CV) that allows the binning of molecular conformations with high dimensionality onto a single coordinate. The choice of a CV is a crucial step in connecting a molecular dynamics simulation to experiments [114, 20]. I come back to the relevance of CVs in the methods part of this thesis and in paper III. The representation in figure 3.1 A shows the rugged, funnel-shaped energy landscape of LacI folding and binding to DNA. LacI is denoted by R in all equations and figures that follow. The specific complex R-O has the lowest free energy where the reduced entropy is compensated for by the formation of protein-protein and protein-DNA contacts. I include the whole process of LacI folding and binding in the representation because I consider the specific complex between LacI and the operator as the fully folded state.

The unfolded protein has high entropy. This entropy is quickly reduced during the initial hydrophobic collapse on time scales of several tens of ns [140]. The formation of α-helices (100-200 ns [179]) and beta sheets (μs) [113] is the next folding step. The system has now reached the Molten Globule state. The formation of further contacts is in balance with the reduction in entropy [118]. This state is characterized by transient interactions within the protein itself where loops might fold and unfold. This feature has been called the fuzziness of protein structure. It is crucial for life as it makes proteins receptive to modulation by their environment [45, 186, 55].

In figure 3.1 A R–N stands for the transient non-specific complex between the partially unfolded LacI and non-specific DNA. As there will not be any formation of a specific complex between LacI and non-specific DNA, the energy landscape is altered (orange in Fig. 3.1 A and B). This should demonstrate that only the interaction with specific sequences will alter the structure of LacI in a way that makes transferring into a specific complex possible.
Figure 3.1. The Energy landscapes of folding, recognition and molecular metamorphosis of LacI (R) in interaction with non-specific (N) and specific DNA (O). Transient complexes are depicted by (---) between the interacting molecules, specific ones by (-). A) The folding and recognition funnel based on Wolynes and co-authors description of folding by the formation of native contacts $Q$ [184]. It describes the gradual reduction in conformational space by the formation of different types of interactions during folding and binding. The alternative energy landscape in orange demonstrates that the non-specific interaction cannot result in the formation of a stable complex. B) Energy diagram of the transition from non-specific to specific binding for the encounter complex with the specific DNA (black) and the non-specific DNA (orange). C) 2D depiction of the different conformational states that can be available to LacI when free (blue) in transient interaction with non-specific DNA (orange, R-N) or in interaction with the operators O1, O2 and O3 which can transition into the specific complex. Here, the same protein sequence can attain different structural forms to perform different functions, just like a butterfly during different stages of its life. Icons created with BioRender.com

Sometimes contacts that are not native to the final structure form during the folding and recognition processes. Non-native intermediates, local minima in the energy landscape, cause the overall folding or binding rate to be slower [33] (see Fig. 3.1 A). This feature is called frustration [184]. Biological systems evade frustration as much as possible by intermediates that are partially disordered so that they can quickly escape local minima. This mechanism, known as fly-casting, is especially relevant for bio-molecular recognition [150, 90]. An important player in stabilizing transient folding intermediates that steer the system towards native folds and specific binding are electrostatic interactions [89, 5]. With DNA being a charged molecule, we can see how non-specific, transient interactions through electrostatics are often favoured when DBPs bind to DNA. The discussion in Chapter 4 will explain why this is crucial for the observed search kinetics of LacI.

Energy diagrams (1-dimensional slices of energy landscapes along one CV), that include thermodynamic and kinetic information for a reaction, can be constructed from carefully run molecular dynamics simulations. MD can sample meta-stable states and conformational re-arrangements between those
metastable states [10, 126]. The energy along the CV (native contacts Q in this case) is obtained from the probability \( P \) to sample a certain CV value. Figure 3.1 B shows two alternative energy diagrams for the transition between the encounter complex between LacI and specific (O) or non-specific (N) DNA R–O/N. The relative stability of the end states is given by the difference in free energies \( \Delta G \) (only shown for the specific binding in the figure). The rate constants of transfer \( k \rightarrow \), the on-rate, and \( k \leftarrow \), the off-rate, can be obtained from the flux between states or from transition state theory via equation 3.2.1 from the height of the transition barrier \( \Delta G^{\ddagger} \) [63, 165, 124].

\[
\Delta G = RT \ln \frac{P(LacI - O)}{P(LacI - O)}
\]  

(3.3)

\[
k \sim \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right)
\]  

(3.4)

Obtaining energy diagrams for a number of encounter complexes R–O/N shown in 3.1 B to observe the effect of the DNA sequence on the barrier height and deduce the binding kinetics, was on goal of my work. The idea is based on MD simulations of enzymatic reactions. Such simulations could show that the catalytic effect of enzymes originates in the pre-organization of charges in the active site. This pre-organization stabilizes the transition state (TS) [176]. Similarly, the specific DNA sequence could stabilize the transition state in figure 3.1 B.

3.2.2 A Short Introduction to DNA Structure

So far I have mostly talked about the structure and folding of proteins, but for the formation of the specific LacI-operator complex, the formation of specific protein-DNA contacts (Q_{protein-DNA}) is required as well (see Fig. 3.1 C). The question about how a protein recognizes a specific sequence is one of the most interesting and rewarding questions left to answer in molecular biology. So, which type of interactions can a protein form with DNA and how are they influenced by the DNA sequence?

DNA famously forms a double helix [178], which is shown in figure 3.2 A. Its building blocks, the nucleotides (Figs. 3.2 B and C), contain a negatively charged phosphate group, a deoxyribose and one of four aromatic (de-localized \( \pi \)-interactions), planar bases: guanine (G), cytosine(C), thymine (T), and adenine (A). Proteins called polymerases synthesize one DNA strand by reading another. This is possible because bases pair according to Chargaff’s rule as shown in figure 3.2 B and C: adenine with thymine (2 h-bonds) and cytosine with guanine (3 h-bonds) [47]. The base pairs form a ladder in the center of the helix, leaving the negatively charged phosphate backbone on the outside of the DNA strand (Fig. 3.2 A). Due to the asymmetric structure
of the bases, B-DNA, the most common form of DNA, has a shallow wide major groove and a narrow and deep minor groove. Figures 3.2 B and C show that the edges of the base pairs present hydrogen bond donors and acceptors in the DNA grooves. A protein can interact with them via its side chains. This is how hydrogen bond donor and acceptor patterns allow direct read-out of DNA sequences by a protein [92, 2]. The bidentate interaction between arginine and guanine shown in figure 3.2 D, for example, is a strong indicator of specificity [98, 136]. Nonetheless, direct readout alone fails to explain the specificity in protein-DNA interactions. As von Hippel points out in his review, hydrogen bond formation between protein side chains and base edges in the DNA grooves does not explain the affinity between a protein and its specific DNA sequence [172]. DNA shape, electrostatics [146, 138, 137] and the interactions of ions and water molecules with DNA need to be considered [97, 77] within a highly adaptable environment [100, 80, 92, 34].

DNA is not very flexible. This is due to several of its physical properties: the $\pi-\pi$ interaction between the bases, the hydrogen-bond formation between them, and the repulsion of negatively charged phosphates in the backbone. The persistence length of B-DNA is about 400-500 Å (150 bp) [108, 14], but varies with sequence [110]. Complexed by proteins though, DNA can be strongly bent [181, 144, 188], base pairing can be disrupted [152], and bases can be flipped out of the helix entirely [71]. Figure 3.3 A shows the example of the bent, A-form DNA bound by LacI [15] as an example. Bending of DNA during specific binding need to be compensated for by an extensive formation of contacts between the binding protein and DNA or by an increase in entropy due to released water molecules from the DNA grooves.
Figure 3.3. Indirect or shape read-out: A) Ball and stick representation of A-DNA. B) Electrostatic potential maps of A- and B-DNA as well as the LacI DNA-binding domain. C) The rotational angles of the deoxyribose. D) Two base pairs with base planes distorted by propeller twists. The molecular graphics have been created in PyMOL [143].

Electrostatic complementary is likely one important driver of DNA bending related to specific binding [30]. Parallels might be drawn to the role of electrostatic pre-organization in enzymes as mentioned earlier [176]. Figure 3.3 B shows the electrostatic potential maps for B- and A-form DNA as well as for the LacI binding domain. We find greater negative potential in the major groove for the A-DNA that matches the positively charged interacting DNA-binding domain.

The DNA backbone is charged making it hydrophilic, but in the apolar grooves, water molecules organize, especially in the minor groove. Freeing these water molecules increases the entropy [97, 77, 128]. Opening of the minor groove to form vdW interactions with the protein is commonly observed in specific protein-DNA complexes [138] including that of LacI [144]. In fact, specific binding of LacI to DNA is entropically driven [132].

Ions likely play an important role in the recognition process as well. On one hand, they influence non-specific interactions by charge screening and they can significantly influence the apparent binding rate of a protein to DNA [12]. On the other hand, the sequence-dependent accumulation of ions on DNA can facilitate sequence-dependent recognition because it influences the electrostatics around the DNA [122, 127, 170, 121].

Finally, something needs to be said about quantifying nuanced changes in the DNA structure due to sequence variability [137]. The parameters that one can check are the torsional angles of the ribose shown in figure 3.3 C or base/basepair parameters. These parameters are related to the relative orientation of the base planes (Fig. 3.3 D) and are explained in detail in an excellent
review by da Rosa and co-authors [36]. We extensively used these parameters in Paper I to study simulations of free DNA.
4. Protein-DNA Interactions in Context

"Curiouser and curiouser!" cried Alice.
From Alice in Wonderland by Lewis Carroll

In chapter 3, I discussed the step-wise folding of flexible polypeptide chains into a rigid specific complex with its operator. In relation to how long the formation of the specific complex takes, the steps up until the formation of the non-specific complex are fast and I will not discuss them here. Instead, I focus on the transient interactions with the millions of non-specific DNA sites that significantly influence the observed binding kinetics. In the following chapter, I discuss how facilitated diffusion was applied to obtain models that explain the observed binding kinetics between LacI and specific DNA. Then, I will introduce the two conformational states of LacI that explain fast DNA search and strong specific binding. Finally, I will set protein-DNA interactions in the broader context of genetic regulation and introduce the model system LacI in more detail.

4.1 Facilitated Diffusion

Let us consider the one-step diffusion-controlled reaction of repressor \( R \) with operator \( O \) in equation 4.1. If we use Smoluchowski’s formalism for diffusion-controlled reactions to estimate the binding rate [155], we picture a specific sequence (\( O \)) as a sink that absorbs a binding protein. The binding rate \( k_a \) in equation 4.2 depends on the following variables: \( D_R \) and \( D_O \) in \( \text{cm}^2\cdot\text{s}^{-1} \), the diffusion rates of the repressor (\( R \)) and operator (\( R \)), respectively, and \( b \), the interaction radius (in cm). \( \kappa \) and \( f_{\text{elec}} \) are the steric and the electrostatic factors [39, 173]. \( N_0 \) is Avogadro’s number. The diffusion coefficients \( D \) can be obtained from the viscosity of the medium \( \eta \) and the size of the particle (\( r \)) according to the Stokes-Einstein relation (Eq. 4.3). We see that, depending on the assumptions made about the shape, charge, and size of the interacting particles, the calculated binding rate \( k_a \) is in the orders of \( 10^7 - 10^8 \text{ M}^{-1} \). I mentioned before that scientists were surprised when they first measured the binding kinetics between LacI and its operator DNA as \( 7 \cdot 10^9 \text{M}^{-1}\text{s}^{-1} \) [133] (Tab. 4.1), which is an order of magnitude faster than estimated in 4.2.

\[
R + O \rightarrow RO
\]  
(4.1)
\[
\kappa_{\text{theo}} = 4\pi \kappa f_{\text{elec}} (D_R + D_O) N_0 / 1000 \approx 10^7 - 10^8 \text{ M}^{-1} \text{s}^{-1} \quad (4.2)
\]

\[
D = \frac{k_B T}{6\pi \eta r} \quad (4.3)
\]

To come up with a model explaining the fast kinetics observed, DNA is considered a long, tube-shaped, negatively charged polymer. The idea is that the protein moves along non-specific sequences of DNA, which reduces the search problem from one in 3 dimensions to one in 2 dimensions. This idea was discussed already before Riggs published the first results on LacI binding kinetics [1].

As Berg and co-authors point out, two steps should be considered to obtain a theory that explains the observed binding rate constants shown in equation 4.4. The authors consider the non-specific interaction of the repressor \( R \) with DNA \( N \) as an intermediate step [19], denoted as \( R-N \) in equation 4.5. This intermediate forms by an electrostatically-driven binding reaction with rate constant \( k_1 \) with any DNA site [93, 40, 131]. A transfer reaction with rate constant \( k_2 \) describes the kinetics of the processes that lead to the formation of the specific complex \( R-O \). The transfer rate includes the rates of several mechanisms via which the protein moves along the DNA chain, such as dissociation and re-association within the DNA so-called DNA domain and 1D diffusion with a diffusion coefficient \( D_1 \) (Fig. 4.1 A).

\[
R + O \xrightleftharpoons[k_d]{k_a} R - O \quad (4.4)
\]

\[
R + N \xrightleftharpoons[k_{-1}]{k_1} R - N \xrightleftharpoons[k_{-2}]{k_2} R - O \quad (4.5)
\]

Firstly, I will consider the equilibrium binding given a two-step binding reaction. The equilibrium constant \( K_{R-N} \) for the formation of the non-specific complex times the total concentration of non-specific sites \( N_T \) will determine the ratio between free and non-specifically bound repressor molecules (Equ. 4.6).

\[
D_IC_{R-N} = \frac{[R-N]}{[R]} = \frac{k_1 N_T}{k_{-1}} \quad (4.6)
\]

The equilibrium binding constant between the repressor and the specific sequence \( K_{R-O} \) is a direct consequence of the stability of the non-specific complex and the transfer rates \( k_2 \) and \( k_{-2} \) as shown in equation 4.7 [19].

\[
K_{R-O} = \frac{[R-O]}{[R][O]} = \frac{k_2 K_{R-N}}{k_{-2}} \quad (4.7)
\]

Assuming that the dissociation rate \( k_{-2} \) from the specific site is small compared to \( k_2 \) and with \( M = \frac{N_T}{O_T} \) being the number of binding sites per chain,
Figure 4.1. The Facilitated Diffusion Model A. A) The DNA domain with the different kinetic parameters that describe the diffusion within this domain as well as to, from and on the DNA. B) The macroscopic binding to the DNA and related rate constants as well as modelling parameters. C) The model for recognition during protein sliding considering a stretch of DNA that contains the specific site marked in green.

The rate constant for the two-step binding mechanism is written as shown in equation 4.8 [19].

$$k_a = \left[ (Mk_1)^{-1} + \left( \frac{k_2 N_T K_{R-N}}{1 + N_T K_{R-N}} \right)^{-1} \right]^{-1} \quad (4.8)$$

The term $(Mk_1)$ describes the first association to the DNA. If the transfer was extremely fast, this rate constant would be equal to $k_a$.

The second part of equation 4.8 depends on the non-specific binding. The factor $1 / (1 + N_T K_{R-N})$ expresses the fraction of time that the protein is free. For high salt and/or little DNA, the protein is mostly free and the transfer rate $k_2 N_T K_{R-N}$ determines $k_a$. If the concentration of DNA $D_T$ equals $K_{R-N}$, half of the proteins are bound. The salt concentration in the cell is in the orders of $10^{-2} \text{ M}$ [108] resulting in a $K_{R-N}$ of about $10^4 \text{ M}^{-1}$ (see Tab. 4.1). The concentration of DNA base pairs in the cell is mM meaning that $N_T K_{R-N} \approx 10$ and that most of the LacI molecules will be non-specifically bound in vivo [64].

The different transfer mechanisms during non-specific interaction are shown in Figure 4.2 A. Most important is sliding [99, 104]. Here, the protein follows the major groove of the DNA with the so-called recognition region (see the section on LacI for further details). Sliding is a rather slow process with a diffusion coefficient that has an upper limit in the orders of $10^{-9} \text{ cm}^2 \text{s}^{-1}$. The slow diffusion is attributed to the helical path along the major groove [145] and occasional contact formation to DNA base edges [190, 105]. Sliding can therefore not explain fast search by itself. In addition to sliding, Berg and co-authors describe different dissociation and reassociation events, some of which transport the protein between correlated, others between uncorrelated DNA sites [19]. Figure 4.1 A shows correlated transfers that occur within one DNA domain: microscopic and intra-domain dissociation and re-association.
Table 4.1. 

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Salt</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{R--N}$</td>
<td>$2 \cdot 10^{-6}$ M</td>
<td>0.05 M K⁺</td>
<td>Lin 1975 [94]</td>
</tr>
<tr>
<td>$K_{R-O_1}$</td>
<td>$2 \cdot 10^{-13}$ M</td>
<td>0.05 M K⁺</td>
<td>Winter 1981</td>
</tr>
<tr>
<td>$K_{R--N}$</td>
<td>$3 \cdot 10^{-4}$ M</td>
<td>0.18 M K⁺, 3 mM Mg²⁺</td>
<td>Lin 1975 [94]</td>
</tr>
<tr>
<td>$K_{R-O_1}$</td>
<td>$3 \cdot 10^{-12}$ M</td>
<td>0.18 M K⁺, 3 mM Mg²⁺</td>
<td>Lin 1975 [94]</td>
</tr>
<tr>
<td>$K_{R--N , dAT}$</td>
<td>$10^{-9}$ M</td>
<td>0.03 M K⁺</td>
<td>Lin 1970 [93]</td>
</tr>
<tr>
<td>$k_a$</td>
<td>$7 \cdot 10^9$ M⁻¹ s⁻¹</td>
<td>0.05 M K⁺</td>
<td>Riggs 1970 [132]</td>
</tr>
<tr>
<td>$k_d$</td>
<td>$6 \cdot 10^{-4}$ s⁻¹</td>
<td>0.05 M K⁺</td>
<td>Riggs 1970 [132]</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>200 s⁻¹</td>
<td>0.05 M K⁺</td>
<td>Lin 1975 [94]</td>
</tr>
<tr>
<td>$k_a$</td>
<td>$3.6 \cdot 10^6$ M⁻¹ s⁻¹</td>
<td>in vivo</td>
<td>Du 2019 [44]</td>
</tr>
<tr>
<td>$k_d$</td>
<td>$1.8 \cdot 10^{-3}$ s⁻¹</td>
<td>in vivo</td>
<td>Du 2019 [44]</td>
</tr>
</tbody>
</table>

with rate constants $\lambda$ and $k_i$ as well as $\Lambda$ and $k_{assoc}$, respectively. This process has been called hopping or jumping depending on the distance that the protein travels [60]. Micro-dissociations are defined via the re-condensation of ions on the DNA [19, 41, 61]. If the protein leaves a DNA domain by dissociating further than a radius $R_c$ (Fig. 4.1 A), it moves by 3D diffusion and is connected to the non-specific complex via the rate constants $k_1$ and $k_{-1}$ (Fig. 4.1 B).

Another mechanism of facilitated diffusion, shown in figure 4.2 A, is intersegment transfer. In this mechanism, one part of the the protein detaches from the DNA and associates to another site close in space but far in sequence (uncorrelated). Eventually, the whole protein transfers. Intersegment transfer likely plays a role in cells [99], but I will not further consider it here as it does not contribute significantly to the observed rate constant.

I will now consider the transfer by sliding. In the model, the target site is effectively extended by the sliding distance ($s = \sqrt{D_1/(\Lambda l^2)}$). The transfer rate constant shown in equation 4.9 will depend on the 1-dimensional diffusion coefficient $D_1$ along the DNA chain as well as the length of this chain $L$, the length of the binding site $l$, assumed to be 1 base pair in this model, and the binding and dissociation to and from the chain $k_{assoc}$, and $\Lambda$ (un-facilitated binding rates). In this formulation of the transfer rate, dissociation and reassociation and the competitive effect of non-specific binding are considered [19].

$$O_Tk_2 = \frac{\Lambda}{(\Lambda l^2/D_1)^{1/2} \coth(\Lambda l^2/D_1)^{1/2} - 1} \quad (4.9)$$

Using $K_{R--N} = k_{assoc} \Lambda$, the expression for $k_a$ in equation 4.8, we obtain equation 4.10 for the association rate constant using sliding as a facilitated transfer.
Figure 4.2. Facilitated diffusion and hypothetical energy landscape for the 3 states of LacI binding to DNA. A) The free repressor first binds with rate constant $k_1$ to any stretch of DNA. It then uses a combination of rotation-coupled sliding, rotation-decoupled hopping, inter-segment transfer, and free diffusion to reach the operator. B) Hypothetical free energy landscape for the unbound protein, facilitated diffusion, and recognition with the rates for binding and unbinding from DNA ($k_1$ and $k_{-1}$) and the rates for finding and recognizing the operator ($k_2$ and $k_{-2}$). The different colors and ways of representing LacI in the different stages along the full binding reaction illustrate that its dynamic changes. We will mostly ignore the unbound state in gray and focus on the non-specifically bound protein in orange and the specifically bound protein in cyan: the search and recognition states. The search state encounters a rough energy landscape when it interacts with the DNA by electrostatic interactions and occasionally hydrogen bonds and van der Waals interactions. The recognition state is tightly bound via aligned charge, charge, and hydrophobic interactions as well as specific van der Waals interactions and hydrogen bonds that are important for recognition [172]. Between the two states is an energy barrier that comes from the entropic loss related to specific binding and conformational switching $\Delta G_{S\rightarrow R}$.

Equation 4.10 predicts that the speed up due to sliding depends on the non-specific binding strength ($K_{R\rightarrow N}$) and the concentration of non-specific DNA $M$. According to the model, long non-specific DNA chains ($\approx 50000$ bp) will compete with the specific operator for LacI. The protein will be trapped in DNA domains far from the target site if the charge screening by salt is weak. The rate constant $k_a$ will therefore be low with very low salt and increase with elevated salt concentrations. The binding constant will pass through a maximum at $N_T K_{R\rightarrow N} = 1$ before it drops with increasing salt when non-specific, electrostatic binding and with it the facilitated diffusion in the form of sliding, wears off [19]. This prediction agrees with the bell-shaped curve seen in experiments [12, 182].
The original theory by Berg has been revisited and simplified many times [23, 62, 61, 154, 72, 109, 64, 99, 59]. At the same time, experimental methods have allowed the observation of molecules diffusing along DNA [175, 174, 104] and the tracking of fluorescently labelled proteins have allowed the study of facilitated diffusion \textit{in vitro} and \textit{in vivo} [46, 64]. Bruinsma as well as Slutsky and Mirny pointed out that a combination of 1D diffusion and 3D diffusion that transfers the protein between uncorrelated DNA domains, will optimize the search [23, 154]. It is not specifically defined in any of these models if the 1D diffusion is by hopping, jumping, or sliding or a combination of these processes. Modelling shows that these processes are likely combined to obtain the optimal search speed [104].

Halford and Marko make the point that sliding more than 10 times the length of the target site is inefficient. Sliding needs to be combined with regular hopping and jumping within the DNA domain to be efficient in search [61]. Relatively short sliding lengths and frequent hopping have been observed for LacI [64, 104]. In fact, these processes lead to LacI missing the target site regularly [104].

Target site bypassing was not considered in the original model proposed by Berg. The probability to bind, \( p_{\text{bind}} \), when the target site is met during sliding, is assumed to be 1 in the original model [19]. In fact, there are several cognate operator sites with different binding kinetics, resulting from different binding probabilities [64, 104, 103]. The equation for the binding probability is shown in figure 4.1 C and it will be discussed in more detail in Paper IV.

4.1.1 A Note on Specificity and Non-Specific Binding

Specificity describes the ability of the DNA-binding protein to identify the operator sequence among a large number of non-specific sites. It can be calculated as the difference in equilibrium binding constants of the specific and the non-specific protein-DNA complexes (\( K_{R-O}/K_{R-N} \)) [94]. As table 4.1 shows, salt influences \( K_{R-N} \) stronger than \( K_{R-O} \) leading to higher selectivity at higher salt concentration. This effect is important on one hand because it suggests that non-specific interactions rely more heavily on electrostatic interactions than specific interactions. This is evidenced by the smaller number of counter-ions released from the DNA upon formation of the specific complex [130, 131]. On the other hand, it suggests that the tightness of the non-specific interaction between protein and DNA may affect the selectivity of protein-DNA interactions. Table 4.1 also shows that DNA with a high content of adenine and thymine has a higher affinity for LacI.

All these effects point towards the importance of non-specific interactions \textit{in vivo}. Antennas, sequences within sliding distance from the specific site, may direct the protein towards the target site [28]. Additionally, high affin-
ity regions that can trap a transcription factor outside its target region, may influence the observed binding rate constants in cells [160, 109].

4.2 The Two-State Model

Facilitated diffusion, as presented above, can explain how a protein can utilize the non-specific DNA around the operator to reach its target sequence faster, but the theory also gives rise to new questions. Physicists determined that for the observed search kinetics, the roughness of the non-specific interaction landscape between proteins and DNA cannot be larger than $1-2 k_B T$ [153]. At the specific site, the roughness needs to be considerably higher for a protein to bind strongly (about 20-30 times the estimated roughness for sliding [18]). How is strong specific and weak non-specific binding to DNA achieved by one and the same molecule?

The answer to this question lies in the internal dynamics of proteins that I discussed in Chapter 3. The so-called two-state model, first suggested by Winter, Berg and von Hippel [182], assumes two main conformational states of the DNA-binding protein that are shown in figure for LacI 4.2 B. One interacts loosely with DNA, the search state, and the other interacts tightly with DNA, the recognition state [153, 72, 18, 88].

Panel B of figure 4.2 shows a hypothetical free energy landscape. It shows the unbound state that quickly transitions to the non-specifically bound state or search state. A large barrier separates this state from the specifically bound recognition state of the protein. Figure 4.2 B demonstrates which types of inter-molecular forces play a role in each state: the search is guided by electrostatic interactions, whereas the specific complex is stabilized by alignment of charges, hydrogen bonds, and van der Waals interactions [102, 88]. The rather large barrier towards the specific complex includes the energetic cost of conformational changes denoted as $\Delta G_{S \rightarrow R}$ [88]. It has been shown theoretically that this conformation switch encouraged by the specific site can resolve the speed-stability paradox in protein-DNA recognition [149]. Modification of the barrier for the conformational change can alter the kinetics of specific binding [189] as I have discussed in Chapter 3 and shown in Paper IV.

4.3 Genes and Genetic Regulation - The lac Operon

We are slowly zooming out. I went from atomistic to molecular interactions and now finally arrive at the bigger picture: genetic regulation. Why is it interesting to study and model how molecules, and in particular proteins and DNA, interact?

---

4I disagree with the name recognition state because I think recognition occurs while the protein is in the search state, but, I stick with the classical terminology here.
I have made the point that functionality is not based on structure alone, but also on selective changes in the structure and complex formation. This is especially relevant for processes that need to be regulated. One such process is the expression of genetic information into proteins.

Genetic regulation is the reason that despite having one set of genes, a cell’s phenotype can vary greatly, in eukaryotes often even more than in prokaryotes [69, 86]. The mechanisms that regulate genetic expression, the gene regulatory networks, constantly react to signals from outside and inside the cell [123, 26]. Their misregulation causes disease [86] or cell death [123].

In prokaryotes, operons are clusters of genes that encode proteins responsible for one cellular pathway, for example the transport and metabolism of certain metabolites. Transcriptional repressors regulate operons by binding to their operators, which prevents the polymerase from binding at its starting sequence, the promoter (see Fig. 4.3). Operons, and particularly the one we are studying, are important biotechnological tools that allow the controlled expressions of genes and the study of genetic regulation itself [83].

The lac operon is shown in figure 4.3. It contains three operator sequences, O1, O2, and O3, the promoter, and the structural genes, lacZ, lacY, and lacA [117, 115]. LacI is active as a tetramer [116]. Two LacI monomers form a dimer with a large protein-protein interface between the core domains (see Fig. 4.3) B. The two C-terminal tetramerization helices of two such dimers
form a 4-helix-bundle, resulting in a dimer of dimers [91, 139]. The native 
LacI-DNA complex in vivo therefore consists of a tetramer bound to two of 
the three operator sites. The DNA loops between those two sites as shown in 
figure 4.3 A.

The specific complex of the dimeric LacI and different operator sequences 
are known from crystallography [91, 15, 16]. The structure with the highest 
resolution contains the dimeric LacI with the strongest known, synthetic, 
operator OSymL (Fig. 4.3 B), PDB-ID 1EFA [15, 159]). The two polypeptide 
chains in cyan and rainbow colors reveal typical features of this specific com-
plex: The N-terminal DNA-binding domains made up of helices 1, 2, and 3 as 
well as the hinge helix (Fig. 4.3 C) and the core region with the effector bind-
ing site [17]. The DNA shown as sticks and spheres in figure 4.3 B is bent in 
the center [157]. Another structural detail of the specific complex that is im-
portant for this thesis is highlighted in figure 4.3 D. This figure shows some of 
the specifically interacting side chains that were identified by Kalodimos and 
co-authors when they studied the dynamical changes in the LacI DBD that 
occur upon specific binding. They found residues that interact with the DNA 
backbone, for example Leu6 and Ser21, [80] and residues interacting specifi-
cally in the major groove: Tyr7, Tyr17, Gln18, and Arg22 [79]. Additionally, 
residues in the hinge helix along with His28 considerably change their dynam-
ics upon specific binding [79], suggesting that they play an important role in 
recognition.

4.4 Structures of LacI - The Role of Flexibility

As we can see in the crystal structure of the specific complex in figure 4.3 
B, the hinge regions, located between the DBDs and the core region of the 
protein, form alpha-helices [157]. They interact with the bent, central part of 
the DNA and specifically with a partially opened minor groove [159]. Sev-
eral studies suggest that the hinge regions are unstructured in the non-specific 
complex and fold in contacts with the specific sequence [157, 81, 79, 163]. 
Mechanistically we can imagine that the positioning of the two DBDs on the 
operator, via specific interactions with base-edges, promotes contact forma-
tion in the hinge regions that favor their folding into alpha-helices [159, 158]. 
The formation of the hinge helices consequently leads to DNA bending at the 
central two base pairs when the minor groove opens to accommodate them 
[81]. Therefore, the whole process has been called coupled binding, folding 
and DNA bending [168].

Figure 4.4 A shows a hypothetical free energy landscape based on insights 
from NMR and crystal structures that was targeted with all-atom simulations 
in Papers I, II, and III. The right bubble shows the known stable complex 
that likely lies in a sharp, deep minimum in the free energy landscape. The 
non-specific complex is more difficult to study as it is dynamic, denoted by
different orange cycles on the rugged landscape in figure 4.4 A. One example is the structure with PDB-ID 1OSL [79] shown in the left, orange bubble in figure 4.4 A. The structure seems to prove that indeed, when interacting with non-specific DNA, the hinge helices of LacI are random coils rather than α-helices as in the specific complex. I want to note that the changes made to the protein that were necessary to perform the experiment can have quite an influence on the conformation of the helix that is interacting with the core domain in the full-length transcription factor. Investigating the importance of flexibility of the hinge region during search was therefore one of the central parts of the thesis and was performed by coarse-grained simulations summarized in Paper II.

The hypothesis of conformational switching and recognition is that it occurs via the folding of the hinge helices shown in 4.4 B. After understanding the role of flexibility in search and recognition, I address how recognition and subsequent conformational switching influence the microscopic binding and unbinding as demonstrated in figure 4.4 C with the mutational studies that target the hinge region, which is described in Paper IV.
Figure 4.4. A) Experimental structures of LacI in the non-specific (orange) and specific complexes (cyan) with a hypothetical free energy landscape that connects them. The large barrier is related to the folding of the hinge helix and the differently colored valleys for the specific complex represent the idea of differently stable complexes with different operator sequences. B) A zoom in on the hinge region in both complexes and a conceptualization of the idea that the rate of folding influences the rate of specific binding. C) The hypothetical free energy landscape of search and recognition with the two rate pairs for non-specific binding $k_{on,max}$ and $k_{off,micro}$, as well as $k_{on,micro}$ and $k_{off,macro}$ for the specific binding based on reference [103]. The molecular graphics have been created in PyMOL [143].
5. Methods

Courage is like - it’s a habitus, a habit, a virtue:
you get it by courageous acts.
It’s like you learn to swim by swimming.
You learn courage by couraging.
Quote by Marie Daly, from GROMACS

In chapter 3, I have described the free energy landscape of folding constructed by binning structural conformations along a reaction coordinate and obtaining the free energy from statistical thermodynamics. There are other types of energy landscapes connected to the free energy landscape and one of them is the potential energy landscape (PES). On this surface the least strained, molecular geometries have the lowest potential energy, they lie in the valleys of the PES. The PES is a function of all atomic positions and can be obtained by calculating the electronic structure of the molecules via methods of quantum mechanics (QM) or via molecular mechanics (MM) [22].

Molecular Mechanics MM, which will be described in detail in this chapter, allows the calculation of potential energy by functions from classical mechanics. Here, I will describe all-atom (AA) simulations, where every atom is a separate unit, and coarse-grained (CG) molecular models, where groups of atoms are treated as one unit. The latter is used to study large systems and long time scales [84].

I used molecular dynamics (MD) to sample the phase space of a molecular system, obtaining positions $r$ and velocities $v$ of the system evolving in time. This sampling can then be used to construct free-energy landscapes.

Fluorescence microscopy instead can be used to study dynamics on the single-molecule level. In the studies here we track fluorescent molecules in vitro and in vivo and combine different theoretical and experimental methodologies to understand molecular search and recognition in complex biomolecular systems.

5.1 Molecular Mechanics

The molecular energy can be determined from the atomic coordinates and the electronic structure. The electronic structure can be calculated by solving the Schrödinger equation in case of the hydrogen atom. For larger systems (still
much smaller than a protein), methods such as density-functional-theory [169] offer approximate solutions but are computationally expensive.

With the implementation of Schrödinger’s equation, it became possible to link vibrational frequencies of molecules measured in infra-red and Raman spectroscopy to a PES and it became clear that bonds and angles of the same types could be treated as harmonic oscillators with the same equilibrium lengths and force constants (chemical transferability) [37]. Angular, dihedral and non-bonded potentials, like the Lennard-Jones potential, are often used to model molecules in chemistry and biology. This methodology is called molecular mechanics (MM) [35].

The classical framework requires the identification of characteristic parameters such as equilibrium values and force constants of bonded interactions, as well as the atoms’ radii and charges to calculate non-bonded interactions. These should either reproduce experimentally obtained molecular geometries from crystallography and NMR, or quantum mechanical calculation. The parameters should also reproduce bulk experimental parameters such as densities, heat of evaporation, or diffusion constants [66].

5.1.1 All-Atom Molecular Models

Molecular simulations support pharmaceutical and biotechnological research and development [70]. They can accelerate experimental work, for example in drug discovery [107], by offering insights into the conformational dynamics of protein-ligand complexes [106, 85]. The all-atom simulations presented in this thesis were performed with the AMBER [27] and GROMACS simulation packages [95] and the AMBER ff14SB force field shown in equation 5.1 with detailed explanation for each term in the following paragraphs.

$$V = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{non-bonded}}$$ (5.1)

**Bonds**

Hooke’s law is the most commonly used model for representing the stretching of bonds between two atoms $i$ and $j$ [35].

$$V_{\text{bond}} = \sum_{ij} k_{\text{bond}} (b_{ij} - b_{ij}^0)^2$$ (5.2)

In simulations involving multiple atom types, the equilibrium bond length, denoted as $b_{ij}^0$, and the force constant, denoted as $k_{\text{bond}}$, are defined for each bond between two atoms.

**Angles**

The bending of an angle between two bonds that connect three atoms, denoted as $i$, $j$, and $k$, is defined as follows:

$$V_{\text{angle}} = \sum_{ijk} k_{\text{angle}} (\theta_{ijk} - \theta_{ijk}^0)^2$$ (5.3)
In this potential $\theta_{ij}^0$ is the equilibrium angle, $k_{\text{angle}}$ the force constant.

**Dihedral Angles**

Dihedral angles are defined as the angles between two bonds, connecting atoms $i$ and $j$ and $k$ and $l$, respectively. Upon rotating about bond $jk$, atoms $i$ and $l$ eclipse each other, resulting in conformations of the molecule that have varying levels of energy, with some being higher and others lower [35]. $V_n$ is the rotational barrier posed by the eclipse of atoms $i$ and $l$, $\omega$ denotes their angle and $n$ describes the periodicity of the potential.

$$V_{\text{dihedral}} = \sum_{ijkl} \sum_n \frac{1}{2} V_n \left[ 1 + \cos(n\omega) \right]$$  \hspace{1cm} (5.4)

**Non-bonded Interactions**

Non-bonded interactions are the interactions that occur between all atoms in the system. These interactions are calculated for all possible pairs of atoms, denoted as $i$ and $j$. The simulation algorithm therefore has to sum over potentials between $N-1$ atom pairs. To make this calculation more efficient, non-bonded potentials are only evaluated if the atoms are closer to each other than a specified cut-off distance [129].

$$V_{\text{non-bonded}} = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij}$$

$$\left( \varepsilon_{ij} \left[ \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \right)$$  \hspace{1cm} (5.5)

The first component of the non-bonded potential accounts for the van der Waals force, which is a short-range attractive force described earlier. To accurately model this force, a short-range repulsive component must be introduced so that electron overlap is prevented. A widely used model for the van der Waals force is the Lennard-Jones (LJ) potential which describes the attractive part as a function of the interatomic distance $r_{ij}$. This distance acts in the attractive part with a power of 6 and in the repulsive part with a power of 12. This may differ in other force fields [66]. The potential reaches its minimum when the atoms are separated by a distance of $r_{ij}^0$. The distance at which the attractive and repulsive forces cancel out is denoted as $r_{ij}$, and at this point, the interatomic potential becomes zero.

Coulomb’s law is used to calculate the interaction between point charges $q_i$ and $q_j$ that are separated by a distance of $r_{ij}$ [50]. The relative permittivity, denoted as $\varepsilon$, represents the ability of the medium between the two charges to screen electrostatic interactions.

Long-range interactions require special treatments in molecular simulations because, on one hand, we are simulating the protein in a periodic box, where
it should not feel its images from the neighboring box. On the other hand, the character of long-range interactions would require the computationally expensive calculation of many pair interactions that scale with the square of the number of atoms [54, 177, 22]. There are methods that allow the reduction of computational costs for long-range interactions such as the Particle-Mesh-Ewald (PME) method [49, 22].
5.1.2 Coarse-Grained Molecular Models

In contrast to the all-atom model, the coarse-grained (CG) model does not treat atoms explicitly. Instead, beads replace groups of atoms [84]. Bonds, angles, and dihedrals that define the geometry of the simplified molecular model are parameterized based on experimental structures. The here used native structure-based Go-Model consists of a bonded term, angle and dihedral terms, a Lennard-Jones-like term, an excluded volume term, and a term describing electrostatic interactions. The force field is shown in equation 5.6. Further details are presented in references [56, 5, 101].

\[ V = V_{\text{Bond}} + V_{\text{Angle}} + V_{\text{Dihedral}} + V_{\text{NC}} + V_{\text{EV}} + V_{\text{DH}} \]  

(Bonded, angle and dihedral potential)

Simple harmonic potentials with equilibrium values \( (x_{ij(kl)}^0) \) based on the reference structure are used to model bonds, angles and dihedrals.

\[ V_{\text{bond}} = \sum_{ij} k_{\text{bond}} (b_{ij} - b_{ij}^0)^2, \quad k_{\text{bond}} = 100 \text{ kcal mol}^{-1} \text{ A}^{-2} \]  

(5.7)

\[ V_{\text{angle}} = \sum_{ijk} k_{\text{angle}} (\theta_{ijk} - \theta_{ijk}^0)^2, \quad k_{\text{angle}} = 20 \text{ kcal mol}^{-1} \]  

(5.8)

\[ V_{\text{dihedral}} = \sum_{ijkl} k_{\text{dihedral}} (\phi_{ijkl} - \phi_{ijkl}^0)^2, \quad k_{\text{dihedral}} = 1 \text{ kcal mol}^{-1} \]  

(5.9)

Native Contacts and Non-native Contacts

The core of the native structure based model is, as the name indicates, its reliance on an experimental reference structure that defines \( A_{ij} \), the equilibrium distance between two atoms in contact. A reasonable force constant \( k_{\text{NC}} \) is assigned to agree with experimental parameters such as the melting temperature.

\[ V_{\text{NC}} = k_{\text{NC}} \left[ 5 \left( \frac{A_{ij}}{r_{ij}} \right)^{12} - 6 \left( \frac{A_{ij}}{r_{ij}} \right)^{10} \right] \]  

(5.10)

To prevent overlap between two atoms, an excluded volume (EV) term with force constant \( k_{\text{EV}} \) is used. \( \sigma_{ij} \) is the summed repulsion radius of interacting partners.

\[ V_{\text{EV}} = \sum_{i<j-3} k_{\text{EV}} (\sigma_{ij}/r_{ij})^{12} \]  

(5.11)
Electrostatic Potential

Partial charges are not assigned in the coarse-grained model, formal charges are. The choices about what to include in a CG model and what not are somewhat arbitrary but should always be guided by the research question. We aim at understanding the interactions between a DNA binding domain, with positive charges, and the negatively charged DNA during facilitated diffusion. This process heavily relies on electrostatic interactions and charges therefore needed to be included. Of course, their interactions are normally influenced by the solvent molecules, but neither ions nor water molecules are explicitly included in the simulation.

The effect of ions and water molecules is modelled by a continuum model instead. The Poisson-Boltzmann (PB) theory can be applied to this problem and further simplified into the Debye-Hückel equation (Eq. 5.12). This is if the electrostatic interaction energy between point charges $q_i$ and $q_j$ is much smaller than the thermal fluctuation $k_B T$ [141]. This is the case when the solution contains only monovalent ions at low concentrations.

\[ V_{DH} = K_{\text{Coulomb}} B(\kappa) \sum_{ij} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}} \]  

(5.12)

Coulomb’s force constant $K_{\text{Coulomb}}$ is approximated in equation 5.13. $\varepsilon_0$ is the vacuum permittivity.

\[ K_{\text{Coulomb}} = \frac{1}{4\pi \varepsilon_0} = 332 \text{ kcal mol}^{-1} \text{Å}^2 \]  

(5.13)

$B(\kappa)$ represents the ion dependency and is approximated to one for dilute solutions [5]. $\kappa^{-1}$ is the Debye screening factor. An expression for $k^2$ can be obtained from the linear approximation of the PB equation and from there we can approximate $\kappa$ as in equation 5.14 [56]. The ionic strength $C_s$ for a solution with only monovalent ions dependents on the concentrations $c_i$ of ions $i$ as shown in equation 5.14.

\[ \kappa \approx 0.32 \sqrt{C_s} \text{Å}^{-1}, C_s = 0.5 \sum_{i=1}^{n_i} c_i q_i^2 \]  

(5.14)
5.2 Molecular Dynamics

A molecular dynamic simulation integrates Newton’s second law of motion in steps of time $\delta t$. Newton’s second law is very powerful because knowing the force $F$ acting on a particle along with its current positions and velocities allows the calculation of its future position. The force is calculated from the potential energy function $V$ in equation 5.15 that can obtained from molecular mechanics [35].

$$ F = -\nabla V = ma = m\frac{dv}{dt} \quad (5.15) $$

An alternative approach to reproduce the thermodynamic characteristics of a system is Monte Carlo (MC) sampling [22], but dynamics will be absent using this method.

There are a number of algorithms that produce trajectories in phase space by numerical integration of the equation of motion. An example of a commonly used algorithm is the Verlet algorithm or modifications of it.

The leapfrog algorithm is a variation on Verlet that evaluates the velocities at half a time-step. The advantage of evaluating the velocities with this offset is that the integration has a smaller error when long time-steps are used. This is ultimately what we would like to achieve for good performance in a MD simulation [53]. The position after a time step $\delta t$ can be obtained as shown in equation 5.16. The velocity is obtained according to equation 5.16 where $a(t)$ is the acceleration and $m$ the mass of the particle.

$$ r(t + \delta t) = r(t) + \delta tv(t + \frac{1}{2}\delta t) \quad (5.16) $$

$$ v(t + \frac{1}{2}\delta t) = v(t - \frac{1}{2}\delta t) + \delta t \frac{F}{m} \quad (5.17) $$

The step size of the MD integrator is an important parameter. It should be as long as possible to make the simulation run fast, but short enough to prevent the overlap of atoms’ van der Waals radii. Different molecular models require different time steps. The time step should be chosen an order of magnitude above the fastest molecular motion in the simulation [22]. Bonds to hydrogen atoms (10 fs) are therefore frozen at their equilibrium length. A common time step in all-atom simulations is 2 fs.

Numerical integration of Newton’s law of motion produces a trajectory in phase space that depends on time. Assuming ergodicity and sufficient sampling over the full phase space, we can calculate ensemble averages for a molecular system sampled in MD simulations via equation 5.18 [35].

$$ \langle X \rangle_{\text{time}} = \frac{1}{M} \sum_{i=1}^{M} X(t_i) \quad (5.18) $$
5.2.1 Modelling Diffusive Processes

The CG-model used in this work does not explicitly include solvent molecules and ions. Instead, the solvent is represented by a continuum model as discussed earlier. The charge shielding effect of ions is included in the electrostatic potential dependency on the ionic strength $c_i$ (Eq. 5.14). However, hydrodynamic effects also play a role in protein diffusion as the protein is regularly hit by randomly moving particles in the solution, changing its dynamics and driving its diffusion. In a continuum solvent, these collisions need to be implemented in the force evaluation. To account for these collisions, Newton’s equation of motion is modified to the Langevin equation of motion (Eq. 5.19). References for this approach include [141, 101].

\[ F = ma + \gamma mv - R(t) \]  

(5.19)

The Langevin damping constant $\gamma$ describes the frequency of collisions. $\gamma mv$ represents dissipation energy caused by friction between solvent particles and the solvent. Equation 5.19 also contains the probability distribution $R(t)$ of random forces exerted on the solute by the moving solvent particles. $R(t)$ has a zero mean, its variance depends on $\gamma$ [141].
5.3 Enhanced Sampling and Metadynamics

To sample the full phase space of a protein is already challenging but it is further complicated by large energy barriers that separate meta-stable states in bio-molecular systems. These barriers are often associated with conformational changes important to folding and binding the processes that I am interested in. Such so-called rare events are far out of reach for a typical simulation time of nano- to micro-seconds [32]. Enhanced sampling methods offer solutions to this challenge.

Enhanced sampling is a broad topic and numerous schemes have been proposed to achieve it. See for example reference [68] for a summary. Here, I used an approach that adds additional potentials to the force field to explore rare events along a reaction coordinate within realistic MD simulation time of nanoseconds: \textit{metadynamics} [21].

Methods, that add so-called bias potential to the force field are much less expensive. The drawback is that these methods require an apriori knowledge of the slowest degrees of freedom, the process with the highest transition barrier. The bias potential will be a function of this collective variable (CV) that captures the slowest dynamics in the system. Examples of CVs are distances between residues or domains, secondary structure [125], or radius of gyration [10].

The most straightforward approach to applying a bias potential is to construct a number of harmonic potentials along the CV. This method is known as Umbrella Sampling (US) [167]. The problem with US is that the system will be forced into configurations that might not correspond to the real system, especially if not all CVs that are relevant for the description of the system’s dynamics are biased. Metadynamics, an alternative to US, will similarly suffer from CVs that neglect slow degrees of freedom [9], but it will report that the CVs are badly chosen. It is important to note for both methods, that, as the sampling is generated based on the sum of the molecular force field and an additional bias potential, the resulting sampling represents the biased configurational space. It needs to be weighted if free energies should be calculated for the unbiased distribution [25]. Even after re-weighting, the results from these types of simulations should validated against experiments.

How can metadynamics report on bad CVs? In metadynamics, the bias builds up during the simulation. Gaussian-shaped potentials are gradually added with a time stride $\tau$ along the CV $S(R)$, where $R$ denotes the molecular coordinates. This is shown in the small plot in figure 5.3 A. The Gaussian-shaped potentials are defined by the height $\omega$ and the width $\sigma$. They discourage the simulation from exploring CV space that has already been sampled. The metaD potential $V_M(S,t)$ (Eq. 5.20) builds up on the underlying energy landscape as shown in figure 5.3 A. Equation 5.20 is written for $d$ CVs $S$,
Figure 5.1. A) Energy landscape along the collective variable S with 3 minima L, M and N that are gradually filled with a time-dependent metadynamics bias potential \( V(S, t) \). B) Scaled bias height along the simulation time \( t \). C) Sampling of the CV \( S \) during the simulation time \( t \). The colors represent the different basins met during the biased simulation.

where figure 5.3 focuses on only one CV for a simple visualisation.

\[
V_M(S, t) = \sum_{t'=0}^{t} \omega(t') \exp \left[ - \sum_{i=1}^{d} \frac{S_i(R) - S_i(R(t'))^2}{2\sigma_i^2} \right]
\]  

(5.20)

In figure 5.3 A, the simulation starts in basin L. Due to the added potentials (also called hills) the system can overcome the barrier to basin M and finally N. The metadynamics flavour that I use and show here is called well-tempered metadynamics [11]. In this method, \( \omega \) is scaled with respect to how much bias has already been added in a region that is currently explored as shown in equation 5.21 and in figure 5.3 B. The parameter \( \Delta T \) determines the simulation temperature for the the biased region.

\[
\omega(t') = \omega_0 \exp \left[ - \frac{V(S, t)}{\Delta T} \right]
\]  

(5.21)

Figure 5.3 C shows how the system explores the molecular configurations with different values of the CV \( S \). At the end of the simulation, the full CV space is explored. This behaviour indicates that the simulation has converged. If we cannot observe the simulation sampling the whole range of the CV at the end of a run, we are likely missing a relevant degree of freedom. This feature of metaD allows us to evaluate the quality of the CVs. Re-weighting can be used to obtain the unbiased free energy landscape from the simulation but it is even possible to extract free energies directly from the bias potential [11]. In another variation of metadynamics, rates of transitions between metastable states can be obtained directly from the biased sampling. This requires that the transitions are fast compared to the rate at which Gaussians are added so that no bias is added to the transition barrier. This method is called infrequent metadynamics [166].
CVs are not only used to enhance sampling but are crucial for understanding the results from a molecular dynamics simulation, defining metastable states, and obtaining transition rates between those states. Therefore, when using metadynamics, the CVs are part of the result [164]. CVs that describe the relevant dynamics of processes, like recognition, allow us to tinker with molecules and change their behaviour on the cellular level in a predictable way, as will see in the last study of this thesis.
5.4 Combining Simulation and Experiments through Rational Design

The idea of predicting molecular function on a cellular level is fascinating and structural biology is the key to doing so. For a long time, crystallography and NMR were the gold standard of this field, but today artificial intelligence (AI) offers ways to predict protein structure and specific interaction from the protein sequence [78, 6]. The example of the non-specific and specific LacI-DNA complexes obtained through NMR and crystallography, respectively, are shown in figure 4.4 A. The missing piece is the prediction of molecular dynamics during transient interaction to obtain the complete energy landscape shown in the center of the figure. Molecular simulations capture dynamics but are difficult and expensive to run and challenging to analyse. They may not capture all relevant states of a system, especially if many different interactions will influence this dynamics.

I therefore came to appreciate experimental methods. Even if they do not reveal molecular structures and dynamic, we can study molecules in their biological context, the cell. When studying proteins, this can be done by altering the amino-acid sequence. I have used MD to identify point-mutations of LacI with varying flexibility in the hinge helix. I then used fluorescent microscopy to study the ability of wild-type LacI and the mutants to find the operator in the cell.

The idea for the design of the LacI variants is illustrates in figure 4.4. Specific binding was thought to be a consequence of stability in the specific complex as illustrated in figure 4.4 A by differently colored valleys on the right side of the energy diagram. According to the two-state model, LacI switches conformations by folding and unfolding of the hinge helix as shown in figure 4.4 B. The formation of the specific complex in turn is triggered by the a DNA sequence that contains specific bases in specific positions, as shown by Spronk and co-authors [159]. The energy diagram in 4.4 A should therefore be modified to the one shown in figure 4.4 C. Here, stronger operators have lower barriers for the formation of the specific complex. This representation based on microarray-measurements of LacI binding kinetics to different operators [103]. I have therefore targeted the hinge-helix of LacI with point-mutations that should change the propensity to form an α helix.

I have discussed the relevance of the search process for the observed binding kinetics of a TF in the cell in Chapter 4. Non-specific interactions will significantly influence binding kinetics on the cellular level. I have therefore used live-cell experiments to understand how the likelihood of helix formation in the hinge helix of LacI influences the in vivo search kinetics.

For this experiment, the LacI gene on the E.coli genome was fused with the gene for the fluorescent Venus protein at the C-terminus. This fusion-protein can be detected with wide-field fluorescent microscopy. With few LacI molecules in the cell and long exposure time, transient binding of LacI to DNA
will be averaged out and specific binding will appear in the form of fluorescent dots that can be counted [46].

We grow the cells on a microfluidic chip with chambers that hold several hundreds of cells in one plane for imaging. In the chip, we can expose the cells to different media. This way, we can dissociate the TF by adding the inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG) that binds to the inducer binding site of LacI (Fig. 4.3 B) and reduces LacI’s affinity to DNA [13]. Upon switching the media back to IPTG-free media, we record its binding by counting the fluorescent dots that appear in the cell when LacI binds.

With this, we have a full circle of structural modelling and measurements that will allow us to decipher the search and recognition process of LacI from its atomistic composition to its behaviour in the cell.
6. Results

The labor into which a heart poured its whole love-
where will it have its say to excite and inspire, and when?

From *Snow Country* by Yasunari Kawabata

6.1 Correlated Motions as Stabilizers of Specific Interactions (Paper I)

Paper I describes the dynamics of the specific complex between the lac repressor and its strongest operator OSymL and non-specific sequences. All studies were based on the crystal structure with PDB-ID 1EFA shown in cyan in Figure 4.4 A. Additionally, we studied the dynamics of the free protein, after deleting the DNA from the complex, to show its flexibility in solution, and the free DNA to identify dynamical features unique to the specific sequence. This part of the study addresses the following questions:

1. How stable is the specific complex and what contributes to its stability (e.g. hydrogen bonds, electrostatics, and correlated motions)?
2. How do changes in DNA sequence influence the stability of the protein-DNA complex?
3. Can we identify DNA shapes that determine specificity?

The specific complex is remarkably stable. When we first tried to simulate disassembly starting from the crystal structure, we ran up to micro-second-long simulations without seeing the protein dissociate. Interestingly, when simulating the protein without the DNA, the DNA-binding domains of LacI demonstrated a large conformational dynamic. The movement of the DNA-binding domains relative to the core domain that we observed was accompanied by a partial unfolding of the hinge region and a de-correlation of its dynamics from the rest of the DNA-binding domain.

To understand how many specific interactions with the DNA sequence contribute to the stabilization of a certain protein conformation in the specific complex, we mutated the DNA in the complex and ran the same simulations. With the change in sequence, we broke hydrogen bonds between protein side chains and base edges but essentially observed similar stability as reported by root-mean-square deviations. Figure 6.1 shows the occupancy of contacts that the protein side chains form to DNA bases (A) and backbone atoms (B) in the
Figure 6.1. Hydrogen bonds between LacI side chains and A) base edges and B) the DNA backbone of OSymL labelled in black versus two other sequences obtained by mutating the DNA labelled in red. The star marks residues placed in the second monomer of LacI and the other side of the operator sequence, respectively, as we look at a symmetric structure. The molecular graphics have been created in PyMOL [143]. This figure has been previously published in my licentiate thesis [96].

native complex with OSymL (black), and the change in occupancy of contacts formed by the same side chain when the sequence has been randomized (red). Tyr17 and Arg22, for example, form strong sequence-specific h-bonds. Upon DNA mutation, those bonds weaken slightly in the case of Tyr17. Whereas the specific contacts formed by Arg22 essentially break on one side of the complex. Gln18 instead binds stronger to the base edges of the non-specific sequences. Instead of loosening the interaction, breaking specific contacts leads to the formation of new contacts. For backbone contacts, which are not directly affected by base mutations, the redistribution of specific interactions results in a slight weakening of all interactions measured.

Besides h-bonds, electrostatics plays a major role in protein-DNA interactions; a role of shape-dependent electrostatics in recognition was proposed for example in reference [31]. We found that the electrostatic interaction energy is about 11 kcal mol$^{-1}$ for the interaction between LacI and straight DNA and 18 kcal mol$^{-1}$ for the specific complex. The differences in these ener-
gies when comparing different sequences are minimal. In total, we could not explain the large differences in non-specific and specific binding with these simulations. The reason might well be that we actually do not capture relevant structures of non-specific complexes. Neither changing the sequence in the specific complex nor assembling complexes from simulations of free protein and DNA yield complexes that we know to occur naturally. Instead, if we just consider the first case, it is likely that the altered DNA sequence would actually favour a different protein conformation. If this conformation is separated from the complex generated by mutations by a large barrier, we will not sample it during 1 micro-second of normal MD. This complex could have completely different electrostatic interaction energies and contacts. The real challenge is therefore not to measure different contributions to the interaction energy but to actually sample relevant molecular arrangements.

The strength of the intramolecular interactions of a protein can be considered equivalent to its stability. Stability can therefore be observed by calculating the dynamic cross-correlation (DCC) between residues. We found the DCC to be high across the DBD when bound to the specific sequence. Mutations in the DNA lead to a loss of dynamic correlation between the first 3 helices of the DBD, especially the hinge helix (residues 51-60). This shows that the specific complex between the lac operator and the lac repressor is characterized by a strong stabilization of the DNA-binding domain. Even if we do not observe significant changes in the RMSD of the complexes with mutated DNA, the DCCM plots demonstrate a change in the overall protein dynamics when the protein starts interacting with another sequence. This relates back to the idea discussed in Chapter 3 and depicted in figure 3.1 C: biomolecular complexes stabilize conformational ensembles of their interaction partners so that they interact with the appropriate strength and selectivity.

The last simulations of this paper were concerned with the role of DNA conformation. The DNA in the specific complex is bent, which could indicate that a pre-bending of the DNA lowers the barrier for recognition. In fact, we could not observe significant differences in DNA parameters when comparing simulations of DNA with different sequences.
6.2 Protein Sliding and Hopping on DNA (Paper II)

Paper I suggests that intramolecular dynamics play a major role in stabilizing specific interactions. This is an important clue when thinking about the recognition of specific sites versus non-specific sites. I have mentioned the speed-stability paradox in protein-DNA interactions that can theoretically be remedied by assuming two different conformations of one biomolecule. Each of these conformations is stabilized in the presence of the fitting interaction partner.

Paper I also suggests that the free protein exhibits large conformational freedom in the DNA-binding domain that is dramatically reduced upon specific binding. NMR studies show that the non-specifically interacting LacI has an entropy similar to the free protein [79]. This suggests that the DNA-binding domain is similarly mobile during the non-specific interaction as in the free state. Based on the discussions in Chapters 3 and 4, we can assume the search state to be a rather flexible state that is mostly sustained by electrostatic interactions. The remaining questions are:

1. How can we simulate the non-specific protein-DNA interaction (search)?
2. Which regions of LacI change conformation between the non-specific and specific interactions and how do we represent that in a molecular model?
3. How can the non-specific interaction between LacI and the DNA be characterized?
4. What could trigger conformational switching at the specific site?

All-atom simulations did not allow us to sample facilitated diffusion that occurs in the order of seconds and minutes. A coarse-grained model with a very low-level molecular description can reach such time scales. We therefore chose to carry out our next set of simulations with a one-bead-per-residue native-structure based CG model which allowed us to study the transient interactions during facilitated diffusion.

With this, the first question is answered. For the second one, many studies, including paper I, suggest the hinge region as the switching region for the transition between non-specific and specific binding. The studies show that the search state is characterized by an unfolded hinge helix and straight DNA during non-specific interaction [79]. Based on this, we constructed two coarse-grained models for DNA and protein, respectively: LacI with a rigid, helical and a flexible hinge helix as well as a straight B-form DNA, and A-type DNA with a central bend.

With question three, we were not aiming at the purely structural characteristics discussed in the previous paragraph but rather the dynamics of facilitated diffusion. Several mechanisms were suggested: sliding, a rotational movement of the proteins recognition region along the grooves of the DNA; hopping, short dissociation and re-association of the protein from and to the DNA; intersegment transfer, transfer between spatially close DNA segments,
and 3D diffusion. We are interested in modelling facilitated diffusion because experiments have suggested that a combination of sliding and hopping is in fact the main driver of fast kinetics despite it leading the target site being occasionally missed. With these experiments as a reference, we ran simulations with the two LacI models and straight DNA. We show that the flexible hinges are crucial when establishing a tight interaction on the straight DNA during sliding. Additionally, flexibility allows an effective switching between sliding and hopping that is likely important for fast target site search.

To answer the last question, we needed to get some clues about recognition. We were wondering if the DNA needed to bend to be recognized by LacI. A target site potential was introduced to study the encounter of both LacI conformations with straight or bent, specific DNA. The target site potential is based on the native protein-DNA contacts found in the crystal structure. We measured electrostatic and specific interactions in four specific encounter complexes (search conformation of LacI with either A or B-DNA and recognition conformation of LacI with either A- or B-DNA). The results show that the DNA-bending and the formation of the helices are concerted processes. The transition into the specific complex follows the encounter of a LacI conformation with an unstructured hinge helix and an operator in B-form that is stabilized by electrostatic interactions but allows for the formation of some specific contacts (base-readout), which is likely important for recognition. Electrostatics also seem to promote the DNA bending that is necessary for the protein to form all specific contacts with the DNA.
6.3 Enhanced Sampling of Specific Recognition (Paper III)

In this study, we used metadynamics to describe and sample recognition between LacI and the operator along one or a few collective variables(s). Being well acquainted with the experimental literature on large systems is really key to finding good CVs. This literature encompasses far more than the crystal structure of a biomolecule or complex (one stable conformation does tell us very little about a dynamic system) but also NMR studies, cryo-EM, and other experimental results. Often, prior coarse-grained modelling will help.

Structures of both states that we want to transition between are very valuable structural input. But caution must be given to the experimental conditions under which structures have been obtained. Another risk is that intermediates are missed by assuming a linear transition between the two experimental end states.

After this general introduction to finding collective variables, I will introduce why and how I applied enhanced sampling along CVs to LacI and its recognition process. The motivation for this work was experiments suggesting the on- and off-rates at the specific site are coupled [104]. Against the common belief that the off-rate determines complex stability, kinetic measurements with different DNA sequences show that the binding probability determines binding affinity [103].

We were interested in sampling the transition involved in the microscopic binding at the specific site, i.e. the transition from an encounter complex to the specific DNA complex. Obtaining kinetics from such simulations and for different DNA sequences would prove the importance of protein-DNA interactions in the encounter complex. I suggest that these interactions in turn determine the binding probability and dictate (together with the off-rate) specificity.

But the challenges of simulating the process of coupled recognition and conformational switching in the protein and the DNA are numerous and enhanced sampling takes care only of the most obvious. Metadynamics will solve the problem of long time scales but it requires a collective variable or a number of collective variables. Finding these collective variables is the main challenge we face, largely because we have limited knowledge of the highly dynamic non-specific complex.

We initially tried to obtain an ensemble of the non-specific complex by using NMR data imposed on the crystal structure. This approach worked to some extent, but left us doubtful because of the changes made to the protein during the experiment. Most importantly though, we did not obtain an ensemble of the non-specific complex from the MD runs with NMR-restraints that agreed well with experimental data. The simulation rather got trapped in local minima that resulted from incompatibilities between the restraints and the starting conformation.
We then tried the most obvious collective variable namely $\alpha$-helicity, protein-DNA contacts, protein-protein contacts, DNA bending and distances, as well as path collective variables using the contact maps (some are not mentioned in the publication), all of which did not give us satisfactory results but rather similar artifacts as seen for the NMR sampling.

In the methods section, I mentioned that CVs have not only been called an input but rather an output of enhanced sampling efforts and that is the case for this study as well. Or rather, it could have been a result, as we never obtained a working combination of CVs. We published our results and inputs in an article and a database as the current developments in the field using AI for CV refinement [29, 151] will be well suited for the task of sampling binding processes between biomolecules.
6.4 Conformational Switching and Recognition in the lac operon Paper IV

In this study, we used MD to select mutants with altered hinge helix flexibility and determined how that influenced binding stability, selectivity, and in vivo search times. We selected two point-mutations in the LacI hinge helix:, one that makes the helix more stable (V52A) and another that makes the helix more flexible (Q55N) compared to the wild-type LacI (WT). Our hypothesis was that a more flexible hinge region would lead to a lower probability of binding the operator, \( p_{\text{bind}} \), less stable binding, and a faster in vivo \( k_{\text{on}} \). Faster binding on the cellular level can be achieved due to less trapping of the protein on non-specific sites.

We can observe this effect in the experiments with E. coli strains that contain WT, Q55N and V52A LacI with and without an operator site. Where the wild-type shows non-specific binding in the form of detectable dots in the cells with exclusively non-specific DNA, the weak binder (Q55N) does not show any dots. Like WT LacI, the strong binder (V52A) exhibits off-target site binding. Off-target site binding by WT and V52A LacI are also present in the cells that do contain an operator.

We observe that LacI binding is affected by the inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG) in all strains. The dots (non-specific and specific) disappear when IPTG is added to the cells with WT LacI and LacI(Q55N). The stronger binder is an exception as IPTG is ineffective for the specific LacI(V52A)-Operator complexes. IPTG is known to stabilize the non-specifically interacting conformation of LacI and to destabilizes the specifically binding conformation (the recognition conformation) [57]. The stronger binding mutant seems to be destabilized by IPTG only when LacI(V52A) interacts with non-operator DNA. We find the influence of molecular interactions on LacI’s conformational dynamics to be very complex. Different DNA sequences in the specific LacI-DNA complex will affect the outcome of ligand binding to LacI differently. We can see that in experiments with a different allosteric factor as well.

All in all, LacI seems to have found an optimal balance between specificity, stability and sensitivity to allosteric regulation. We can show that we can shift this balance either in the direction of specificity or stability by changing the stability of the hinge region.
7. Conclusions

This thesis aims at understanding the structure and function of transient intermediates which are crucial for molecular recognition in living cells. Despite their biological importance, these intermediates have long been overlooked because they are not readily accessible to structural biology methods [55, 119].

Proteins that are synthesized in one part of the cell need to reach their site of activity without interacting too strongly with other cellular components they might encounter. Additionally, proteins are often regulated by interactions with small molecules, chemical modifications, changes in temperature, pH, or interactions with other macromolecules. Selective and regulated molecular interaction is possible because the environmental conditions that a protein finds itself in, alter the stability of different protein structures with different functions [135]. Their partial flexibility allows proteins to balance transient, non-specific interactions with stable and permanent interactions at the target sites. This is the basis of cellular signalling [186] and, most important for this thesis, transcriptional regulation [161, 82].

I have discussed how difficult it was for us to study encounter complexes of LacI and DNA to understand specificity. Nonetheless, our study gave us many clues as to how different sequences of DNA influence the LacI structure. We find that the DNA sequence is directly related to the correlation of motions in LacI itself. We find that non-specific binding is stabilized by electrostatics and that the intrinsic disorder in the hinge helix region of LacI allows probing of the DNA for base-readout. Base-readout, in turn, is important for recognition as it triggers the switch into the specific binding conformation.

Not only the DNA sequence is important for recognition at the target site, but also the proteins’ likelihood to switch conformation. We see in the in vivo studies how LacI balances the stability of the complex (k_{off,micro}) with the likelihood of specifically binding (also to non-operator sequences) (p_{bind}) and its ability to react to IPTG. This balancing act is a feature of the hinge helix and its intrinsic flexibility. With this knowledge, we can design LacI mutants with changed complex stability and search times as well as sensitivity to different allosteric factors. In the future it will be rewarding to study the interplay between DNA-sequence, allosteric factors and point-mutations (similar to this study [57]) in the hinge-region to understand specificity and binding kinetics and its relation to conformational states. This will be possible by combining molecular dynamics simulations and experiments in vitro, to study microscopic binding kinetics, and in vivo, as shown in Paper IV.


Das Problem mit flexiblen Strukturen ist, dass man sie auch mit modernen Techniken der strukturellen Biologie schlecht bis nicht sichtbar machen kann. Um sie zu studieren, benutze ich Molekulardynamische (MD) Simulationen. Diese basieren auf Kraftfeldern, die auf experimentelle Strukturen angewendet werden können, um ihre Dynamik zu simulieren.

Mit dieser Technik habe ich gelernt, warum eine flexible Region im Transkriptionsfaktor LacI die Region ist, mit der das Protein kurzlebige und stabile Interaktionen mit DNA so ausbalanciert, dass es möglichst schnell an den Ort gelangt, an dem es seine Funktion ausüben kann. Das ist möglich, weil sich diese meist flexible Region, die auch Gelenk-Region genannt wird, erst
dann stabilisiert, wenn der Transkriptionsfaktor mit einer bestimmten DNA-Sequenz interagiert. Durch die Stabilisierung dieser Region ändert sich die Struktur des Transkriptionsfaktors und damit seine Affinität zu DNA.

Mit diesem Wissen ist es möglich, LacI Varianten zu entwickeln, die DNA mit unterschiedlicher Stärke binden, unterschiedlich schnell an den Ort ihrer Aktivität gelangen und auch unterschiedlich auf Moleküle reagieren, die die DNA Interaktion verändern.
Livet är interaktionen mellan molekyler. Cellerna är proppfulla av dem och varje molekyl måste vara på rätt plats vid rätt tidpunkt. För att åstadkomma detta interagerar molekylerna olika länge med varandra. Medan vissa bildar långlivade (minuter till timmar) komplex med sina inteaktionspartners, interagerar de flesta molekylerna bara under en kort tid (små bråkdelar av en sekund).

Proteiner och deoxiribonukleinsyror (DNA) är olika typer av makromolekyler i cellen. Proteinerna är cellens arbetare medan DNAT är informationsbäraren som tillhandahåller ritningarna för hur proteinerna ska sättas ihop. Regleringen av informationsflödet från DNA till proteiner har många nivåer och längst ner hittar vi avläsningen av DNA, transkriptionen. Transkriptionsregleringen utförs av en klass DNA-bindande proteiner, de så kallade transkriptionsfaktorerna. I mitt arbete studerar jag strukturen hos transkriptionsfaktorn LacI och hur proteinet interagerar med DNA.

Jag är särskilt intresserad av hur olika strukturella egenskaper hos LacI möjliggör selektiva kortlivade och långlivade komplex med DNA beroende på sekvens. Proteiner är relativt stora molekyler som består av mindre enheter med olika kemiska egenskaper. Dessa enheter bildar långa kedjer som veckas till tredimensionella strukturer inuti cellens cytoplasma som är fylld med vatten och andra molekyler. Proteinets struktur bestäms av följden av enheter i kedjan. Vissa sekvenser bildar helikala strukturer, medan andra bildar bladliknande flak. De olika tredimensionella strukturerna monteras ihop till så kallade proteindominator, som är ganska stabila. De flesta proteiner har emellertid också flexibla regioner som ofta återfinns mellan proteindominatorna och spelar en viktig roll i bildandet av specifika komplex med andra proteiner, små molekyler eller DNA.

Problemet med de flexibla strukturerna är att de är svåra att observera även med moderna strukturbioologiska tekniker. För att studera dem använder jag mig därför av molekylärodynamiska (MD) simuleringar. Simuleringsarna baseras på kraftfält som kan appliceras på experimentella strukturer för att få en uppfattning om deras dynamik. Med hjälp av MD har jag lärt mig varför en flexibel region i transkriptionsfaktorn LacI är nyckeln till balansen mellan proteinets kortlivade och stabila interaktioner med DNA som gör att LacI snabbt kan närstå plats i cellen. Denna flexibla region, också kallad gångjärnsregionen, stabiliseras nämligen först när transkriptionsfaktorn interagerar med "rätt" DNA-sekvens.

Genom att ändra stabiliteten hos gångjärnsregionen kan vi modifiera transkriptionsfaktorns struktur och därmed dess affinitet för DNA. Med denna
kunskap är det möjligt att utveckla LacI-varianter som binder till DNA med olika styrka, när sitt mål olika snabbt och som svarar olika på andra regulatoriska molekyler.
First and foremost, it is Johan Elf who I want to thank. You helped me to continue at times when I did not believe in myself anymore. Joining your group gave my studies and my life a whole new, positive direction and I am immensely grateful for that.

I am also grateful to my collaborators: Yaakov Levy and his group, Lavi, Arumay and Harry, at the Weizmann Institute of Science in Israel. I spent 6 months filled with great science, discussions and continuous support there; and Gareth Tribello at Queen’s University in Belfast, where I spent the most rewarding and productive week of my life. I also want to thank my co-supervisor David Van der Spoel for discussions, help and input.

I also want to thank Lynn Kamerlin and the people I got to work with in her group. Paul, Miha and Anna taught be the basics of molecular simulations. I want to thank Dennis and Qinqua, who worked with me on the LacI project, and Yash, Klaudia, Dusan, Jasmine, Marina, Rita, Catia, Rory, Michal and Andrey for many refreshing moments during lunch. I also want to thank my former co-supervisor Mikael Widersten for his support.

I have the most wonderful colleagues. I look forward to work everyday. I want to thank Irmeli for your skill and help, when things seem impossible, and for the editing of this thesis (any mistakes were introduced by last minute changes). Magnus, for giving me some good advice. Jimmy, thank you for your wisdom and for your honesty, that is a real treasure, for perfect curves and for a working lab. Spartak, for fixing my code, teaching me a little matlab, always being enthusiastic about running and skiing trips. David, for the inspiring discussions about data management and cellular diffusion (cells are tiny and things move very fast). And for reading and improving this thesis. Elias, for being kind, patient, curious and always helpful. Anna, for teaching me the secrets of microbiology. Beer, for being a great partner in that business now. Camsund, thanks for sharing your insights on academia and the world outside. Jones, thank you for answering my questions about biophysics (and hiking). A special thanks to Johan Öman for fixing the espresso machine, baking buns and generally being cheerful company in the morning.

In the strict lab hierarchy, lore-master, dungeon keeper and the inner circle are followed by the senior post-doc. When I joined, that was Kuba. Thank you for the songs, that made my dark days feel a little brighter and for being a true friend. I miss you.
Now, Vinodh is the senior-postdoc and you, too, cheered me up many times, thank you. I want to thank Dvir for being interested, open-minded and a passionate scientist. Daniela for always being happy and introducing us to Mexican snack (and drinks). Bo, it was fun with you.

Praneeth, you will finish and move on to build the life you want for yourself. Thanks for helping me do the same. Don’t get lost! Space is vast...

Konrad, thanks for making coffee for all of us and for finally inviting us over (I know this hasn’t happened, but...). Oscar, thank you for your calmness, humour and for solving the mysteries of replication initiation. Jinwen, Thanks for being my LacI sister. I am excited to see how far you will bring the little transcription factor in your PhD. Minh, thanks for turning the office back into a place of tunes and happy plants. Henar, thank your for all the energy you bring to our lab. Danna, thanks for giving me so much time on the scope when you also needed it. I really appreciated that. Filip, it is good to have someone around to connect with over exits, books and really, really bad jokes. Tora, thanks for being a strong and kind person who inspires.

Javier for listening and the luxury it is to hang out with you. Laura, thanks for asking me out, for finding the right words and taking me on the most wonderful trip.

Finally, thanks to all the members of the Elf, Johansson and Deindl groups, for creating a fun and friendly working environment!

Emil, despite having been a massive help and important discussion partner on the last project included in this thesis, you are probably the person in the lab I have known the longest. We had our disagreements but in the end, you have always been a friend and I am very grateful for that.

The MD people: Erik Marklund and his lab in Chemistry, David VdS’s group, especially Alfred, who was always enthusiastic about our little journal club, along with people in Johan Åqvist’s and Jens Carlsson’s group with a special shout-out to Lucien, Florian and Alejandro. Good luck to all of you!

Filip and Elin, you are my oldest friends in Uppsala and so difficult to stay in touch with: It is always great to catch up with you! I hope we will continue doing it for the years to come!

Giulia, thank you for becoming my friend, because I enjoy German poetry (which you found ridiculous). I treasure the times we find to catch up and share thoughts. Joao for being a wonderful person. Joana, for showing me a deep passion for creating beauty and thinking freely. Fabian, for your kindness (and that includes your family) and a wonderful trip to the Alps. Johanna, thank you for the runs in the forest and for our lunch conversations.

Nitai and Sarah, thank you both for giving me a home far away from home. And Sarah, for that assurance that we can never ask for, or be, too much.

The Real deal! You have accepted me as I am and given me the happiest and saddest times. Caro, thanks for making me search for my values, your love for poetry and music; Boris, for inspiring me to write, always being honest and with an open heart; both of your for giving me a home when I needed one;

60
Elena, for going with me through some difficult moments and believing in me; Julian for showing me how chill things can be taken; and Sascha for being a free spirit and loyal friend. I hope we will continue just like we were. I miss you. Also, thanks, Yunus, for checking in from time to time.

Catharina, Eike and Lydia (Team Tequila), for the good times in Leipzig and the meetings we had every year since most of us moved on to new places. Adrian, Silja, Finja und Tarje (and now Hilja and Rune, who I am excited to meet) for being welcoming and warm.

I want to thank the people from back home who supported me from far: Regina, even if you disappear, you’re back soon enough, and it is all the same. Katharina, thanks for being around in 2018, the letters and calls. Januni, for texting from all over the world. Annika, thank you for giving me advice and insights.

Elias, you are not in my life anymore but, you and your family have influenced and I am very grateful for this part of myself. You will always inspire who I am and what I wish for.

My great-grandmothers Hilda and Martha for sharing their memories, the history of our families and for the right amount of restlessness.

Ilka, für die Gewissheit, dass am Ende alles gut wird, wenn wir in uns selber ruhen. Nils and Fynn, egal, was passiert, solange ihr glücklich seid, bin ich glücklich. Ich bin stolz auf euch beide. Oma Helma für die Einblicke in ein Leben, dass lange vergangen ist und doch alles beeinflusst, was heute meine Realität ist. Oma Liane and Opa Werner, für die Tage im Garten, für Geschichten, Stil und Werte. Papa, für den Blick in die Natur, der ihre Schönheit offenbart. Mama, für meinen eigenen Weg, den du mich hast finden lassen.
References


[130] M. T. Record, Jr, P. L. deHaseth, and T. M. Lohman. Interpretation of


[148] Y. Shi. A glimpse of structural biology through x-ray crystallography. Cell,


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)