Modelling Protein Dynamics on the Microsecond Time Scale

Problems and Solutions

Ph. D. Dissertation
Iwona Anna Siuda, 20097241

28th January 2013

Supervisor: Christian N. S. Pedersen
Co-supervisor: Lea Thøgersen

Bioinformatics Research Centre
Department of Molecular Biology and Genetics
Aarhus University
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>I</td>
</tr>
<tr>
<td>Preface</td>
<td>V</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>VII</td>
</tr>
<tr>
<td>Abstract</td>
<td>IX</td>
</tr>
<tr>
<td>Dansk Resumé [Danish Summary]</td>
<td>XI</td>
</tr>
</tbody>
</table>

1 General Introduction  
1.1 What is a Protein? .................................. 2  
1.2 Protein Structure .................................. 2  
1.3 Malfunction of Proteins ............................ 2  
1.4 Experimental Structure Determination .............. 3  
  1.4.1 X-ray Crystallography .......................... 3  
  1.4.2 Nuclear Magnetic Resonance ..................... 4  
1.5 Computational Modelling ............................ 4  
  1.5.1 Force Field .................................... 5  
  1.5.2 Molecular Dynamics .............................. 6  
  1.5.3 Atomistic Resolution ............................ 7  
  1.5.4 Coarse-Grained Resolution ...................... 7  
  1.5.5 Standard MARTINI Model ......................... 8  

2 Studying Protein Domain Movements - domELNEDIN Model  
2.1 Protein Domains .................................... 12  
2.2 domELNEDIN Model ................................... 13  
2.3 Test Example - Periplasmic Leucine-Binding Protein  
  ..................................................... 14  
  2.3.1 LBP Domain Definitions .......................... 15  
2.4 Methods ............................................ 16  
  2.4.1 LBP Setups ..................................... 17  
  2.4.2 Simulation Details ............................... 17  
2.5 Protein Dynamics on the Nanosecond Time Scale .... 17  
  2.5.1 Stability of Protein Structure in the MARTINI Model  
       .................................................. 17  
  2.5.2 ELNEDIN Scaffold ................................. 18  
  2.5.3 domELNEDIN Scaffold ............................. 24
## 2.6 Protein Dynamics on the Microsecond Time Scale

- 2.6.1 Starting from the Open Conformation ........................................ 25
- 2.6.2 Starting from the Closed Conformation ........................................ 28
- 2.6.3 Protein-Ligand Interactions ..................................................... 31

## 2.7 Conclusions ................................................................................. 33

## 3 Improving CG Modelling of Protein-Ligand Interactions

- 3.1 Polarizable MARTINI Force Field ................................................. 36
- 3.2 Polarizable Water Model ............................................................. 38
- 3.3 Methods ....................................................................................... 39
  - 3.3.1 Software ................................................................................. 39
  - 3.3.2 LBP Setups ............................................................................. 39
  - 3.3.3 Protocol .................................................................................. 39
- 3.4 General Corrections to Standard MARTINI Model ...................... 40
- 3.5 Improving Protein-Ligand Interactions ........................................ 41
  - 3.5.1 Leucine Ligand Description for Polarizable Model .................. 41
  - 3.5.2 Closed Conformation of LBP .................................................. 42
    - Removing Ligand from the Binding Cleft .................................... 44
    - Pulling Domains Apart ............................................................... 45
  - 3.5.3 Open Conformation of LBP .................................................... 46
    - Random Positioning of Ligands .................................................. 46
    - Forming a Salt Bridge ............................................................... 46
- 3.6 Conclusions ................................................................................. 49

## 4 SERCA and Cholesterol

- 4.1 Ca\(^{2+}\)-ATPase (SERCA) .......................................................... 52
  - 4.1.1 The SERCA Membrane Environment ...................................... 54
  - 4.1.2 Annular vs. Non-annular Binding Sites ................................... 54
  - 4.1.3 Cholesterol as a Regulator ..................................................... 54
- 4.2 Methods ....................................................................................... 56
  - 4.2.1 Protocol ................................................................................ 56
  - 4.2.2 Simulation Parameters ........................................................... 57
  - 4.2.3 Analysis ................................................................................. 57
- 4.3 Results and Discussion ............................................................... 58
  - 4.3.1 Bilayer Self-Assembly ............................................................. 58
  - 4.3.2 Transmembrane Surface of SERCA ......................................... 59
  - 4.3.3 Possible Interaction Spots for Cholesterol ............................... 59
  - 4.3.4 Annular Interactions ............................................................... 61
  - 4.3.5 Potential Non-annular Binding Interactions ............................ 63
- 4.4 Perspectives ............................................................................... 67
- 4.5 Conclusions ............................................................................... 68

## 5 Dynamics of Saposin-Like Proteins

- 5.1 Saposins A, B, C, D, and NK-lysin .............................................. 70
  - 5.1.1 Hypothetical Binding Modes to Biological Membranes ............ 72
- 5.2 Methods ....................................................................................... 74
  - 5.2.1 System Setups ....................................................................... 74
  - 5.2.2 Simulation Protocol ............................................................... 74
    - AMBER09SB-ILDN Simulations ............................................... 75
    - GROMOS45a3 Simulations ...................................................... 75
    - MARTINI Simulations ............................................................. 75
  - 5.2.3 Analysis ................................................................................. 75
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.4 Abbreviations</td>
<td>75</td>
</tr>
<tr>
<td>5.3 Structure Changes in AMBERSB99-ILDN and GROMOS45a3 Force Fields</td>
<td>76</td>
</tr>
<tr>
<td>5.4 MARTINI CG Simulations</td>
<td>81</td>
</tr>
<tr>
<td>5.4.1 Water Model Influence on Protein Structure</td>
<td>83</td>
</tr>
<tr>
<td>5.4.2 Biological Implications</td>
<td>83</td>
</tr>
<tr>
<td>5.5 Conclusions</td>
<td>84</td>
</tr>
<tr>
<td>6 Conclusions and Future Perspectives</td>
<td>85</td>
</tr>
<tr>
<td>Bibliography</td>
<td>89</td>
</tr>
<tr>
<td>Appendix A to Chapter 2</td>
<td>103</td>
</tr>
<tr>
<td>Appendix B to Chapter 2</td>
<td>157</td>
</tr>
</tbody>
</table>
Preface

The work presented in this thesis has been performed during my Ph.D. studies initiated in November 2009 at the Department of Molecular Biology and Genetics, within the field of bioinformatics at the Bioinformatics Research Centre (BiRC), Aarhus University, Denmark. During this period of time I have been associated with the Centre for Membrane Pumps in Cells and Disease (PUMPKIN), Danish National Research Foundation, Aarhus University. The work has been carried out under the supervision of Associate Professor Christian N. S. Pedersen and Associate Professor Lea Thøgersen.

As a part of my Ph.D., I spent seven months (February 2011 - July 2011, and February 2012) in the laboratory of Professor Peter Tieleman at the University of Calgary, Canada. Three weeks in the summer (July/August 2012) I spent in the laboratory of Professor Siewert-Jan Marrink at the University of Groningen, the Netherlands. The stay was financed by the HPC-EUROPA2 Transnational Access program (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures.

My studies have resulted in one manuscript submitted for publication, and three manuscripts under preparation. The four papers are:

- **I. Siuda, L. Thøgersen, Protein Domain Coarse-Grained Molecular Dynamics: Application to the Leucine Binding Protein.** Submitted to *Journal of Chemical Theory and Computation*. Appendix A.

- **I. Siuda, D. H. de Jong, L. Thøgersen, S.-J. Marrink, Protein-Ligand Interactions of Leucine Binding Protein in MARTINI Polarizable Force Field.**

- **H. E. Autzen, I. Siuda, L. Thøgersen, Regulation of SERCA by Cholesterol - a Direct or Indirect effect?**

- **I. Siuda, S. Baoukina, D. P. Tieleman, Molecular Dynamics Studies on Structural Changes in Saposins A, C, D and NK-lysin.**
Ten posters have been prepared in collaboration with Lea Thøgersen and presented at international conferences and workshops:

- **Coarse-Grained Biomolecular Modelling, CECAM, Coarse-graining workshop**, Lausanne, Switzerland (2011).
- **Biological Sciences Graduate Students’ Annual Meeting**, University of Calgary, Canada (2011).
- **Conference on IT research at AU: Department of Computer Science**, Aarhus University, Denmark (2010).

The posters are not included in this thesis.

The thesis starts with the introductory Chapter 1, which presents the general introduction to proteins and methodology used for their study. The subsequent Chapters 2-5 will cover my contribution to the research projects I was involved in. Finally, the thesis is ended with overall conclusions and future perspectives.
During three years of my Ph.D. I met many people whom I wish to acknowledge for the support and opportunities they gave me.

First, I would like to thank my supervisors Assoc. Prof. Christian N. S. Pedersen and Assoc. Prof. Lea Thøgersen for accepting me in the group, for creating a great work environment and giving me lots of opportunities for traveling to workshops and conferences. Lea, also for being supportive, for inspiring ideas and discussions, for general guidance, for contribution and involvement in the projects described in this dissertation and for giving me feedback on this dissertation.

I had the pleasure of working with Prof. Peter Tieleman. My great thanks to Peter for accepting me twice in the group, for being encouraging and enthusiastic towards my work, for constructive and inspiring discussions, and for letting me spend a great deal of his allocated CPU-time. Furthermore, thank you to Peter and Svetlana Baoukina for collaboration on the saposin-like proteins project.

Thank you to Prof. Siewert-Jan Marrink for allowing me to visit and work in his group. Furthermore, thank you to Siewert-Jan and Djurre H. de Jong for collaboration on the MARTINI protein-ligand project.

Also big thanks to the inspiring and kind people in the groups of both Peter Tieleman and Hans Vogel at the University of Calgary, Canada, and people in the group of Siewert-Jan Marrink at the University of Groningen, The Netherlands.

Many thanks to all the members, past and present, of the Biomodeling group at the Department of Chemistry, University of Aarhus, Denmark. I am thankful for the great atmosphere during the group meetings and social events. Special thanks to Julie Grouleff, Anne Laustsen, Henriette E. Autzen, and Ole J. Andersen for proofreading sections of this dissertation. Henriette, also for collaboration on the SERCA project. Also great thanks to Johan F. Kraft for allowing me to use his \LaTeX{} template, on which this dissertation was assembled.

Many thanks for the people from the Bioinformatics Research Centre at University of Aarhus, Denmark, for a great working environment. Thank you to Ellen Noer for helping
me with many problems concerning moving between countries, and with all administrative issues. Special thanks to Yu Qian, Jinjie Duan with whom I have shared many thoughts, joys and - from time to time - frustrations and sorrows during the last three years. Thank you for many hugs and good words.

Thank you to the Centre for Membrane Pumps in Cells and Disease (PUMPKIN), and the Graduate School of Science and Technology for funding my Ph.D. Also thank you for the support from the Department of Molecular Biology and Genetics for traveling grants. Thank you for funding from the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures, which provided computing allocations at SARA in the Netherlands.

Special thanks to Filip Pawlowski for giving me the opportunity to study abroad and helping me with technical problems with \LaTeX{} during the writing process.

Finally, I wish to thank my parents and my sister for their faith in me and words of support. Mom and Dad I am very lucky to have such wonderful, loving, and supportive parents. Thank you.

Iwona Anna Siuda

Aarhus, 28th January 2013
Abstract

Recent years have shown increase of so-called coarse-grained (CG) molecular dynamics (MD) simulation methods that provide structural and dynamic details of large proteins and self-assembly of biological materials, that cannot be easily probed experimentally and are still limited in atomistic simulations. In this dissertation the dynamical properties of globular and membrane proteins are investigated using MARTINI CG model, as well as an extension to the model, ELNEDIN, which utilizes an elastic network to stabilize the tertiary structure of proteins. Also, an alternative extension to MARTINI, referred to as domELNEDIN, is presented, which allows protein domains to move independently and thus, introduces flexibility that should allow for the simulation of conformational changes.

With MD studies of the periplasmic leucine-binding protein (LBP) in water it is illustrated that MARTINI is unable to ensure structural integrity of a protein with high β-sheet content, while in the ELNEDIN model the structure is strongly biased to the initial structure. It is also presented that the domELNEDIN model stabilizes the structure in the same manner as the ELNEDIN model, while introduces flexibility that may lead to conformational changes. Modelling of the LBP’s interactions with ligand is not possible in the standard MARTINI model, however, a newly presented version of MARTINI model, where the additional charges on polar particles and an “off-centre charge” model for charged residues are introduced, allows the study of ligand effects on the conformational flexibility of LBP.

It is found experimentally that the sarco(endo)plasmic reticulum Ca\textsuperscript{2+}-ATPase (SERCA) interacts with cholesterol in the surrounding membrane, and cholesterol, in some cases, also affect the activity of the protein. MD simulations where SERCA is described using MARTINI and ELNEDIN models, and the lipid and cholesterol molecules self-assembly into a mixed bilayer around the protein are presented and used to derive the potential cholesterol-binding sites on the transmembrane surface of SERCA. Finally, the MD study
of four proteins belonging to the family of saposin-like proteins, using two atomistic and to CG models is presented and discussed.
Detaljeret indblik i store proteiners struktur og funktion er svært at opnå med ren eksperimentelle metoder og er stærkt begrænset i tid og udstrekkning i simuleringer med atomær oplosning. Udvikling og anvendelse af såkaldt grovkornede molekylodynamiske (MD) simuleringstekniker er derfor blevet mere udbredt de seneste år. I denne afhandling er de dynamiske egenskaber af både globulære og membranindlejrede proteiner studeret ved brug af den grovkornede MD model MARTINI, samt en overbygning til modellen, ELNEDIN, som udnytter et elastisk netværk til at stabilisere den tertiære struktur af proteiner. I afhandlingen præsenteres desuden en alternativ overbygning til MARTINI, som ligeledes anvender et elastisk netværk til at stabilisere strukturen, men med en modificering således at proteinet kan ændre konformation. Vi kaller denne alternative elastiske netværksmodel for domELNEDIN.

Ved MD studier af det periplasmiske leucin-bindende protein (LBP) i vand illustreres hvordan MARTINI modellen ikke kan stabilisere den tertiære struktur i sig selv, og at ELNEDIN sørger for at stabilisere strukturen, men samtidig udelukker konformationelle ændringer i strukturen. Desuden ses at domELNEDIN modellen stabiliserer strukturen på samme vis som ELNEDIN, samtidig med at fleksibilitet der kan føre til konformationelle ændringer tillades. Modellering af proteinets interaktion med sin ligand, leucin, er ikke mulig i den oprindelige MARTINI model, men en nyligt præsenteret opgradering af MARTINIs proteinmodel, hvor ladningsfordeling på polære sidekæder kan præsenteres, muliggør også studiet af ligandens effekt på den konformationelle fleksibilitet af LBP.

Det er eksperimentelt vist at den sarco(endo)plasmiske reticulum Ca$^{2+}$-ATPase (SERCA) interagerer med cholesterol i den omgivende membran, og at cholesterol i visse tilfælde også påvirker proteinets aktivitet. MD simuleringer hvor SERCA beskrives i en kombineret MARTINI-ELNEDIN model og et blandet lipid-cholesterol bilag samler sig omkring proteinet præsenteres og anvendes til at udlede potentielle cholesterol-bindingssteder på
SERCAs overflade. Til sidst præsenteres et studie af fire proteiner tilhørende familien af saposin-lignende proteiner, hvor to atomistiske og to grov-kornede modeller sammenlignes og diskuteres.
Chapter 1

General Introduction

This chapter provides a general introduction to proteins and the experimental and computational techniques that can be used for their study. A special emphasis is put on the molecular dynamics simulation method, being the corner stone for the work presented in this dissertation.

Contents

1.1 What is a Protein? .............................................. 2
1.2 Protein Structure .............................................. 2
1.3 Malfunction of Proteins ........................................ 2
1.4 Experimental Structure Determination ......................... 3
  1.4.1 X-ray Crystallography ...................................... 3
  1.4.2 Nuclear Magnetic Resonance .............................. 4
1.5 Computational Modelling ....................................... 4
  1.5.1 Force Field ............................................... 5
  1.5.2 Molecular Dynamics ...................................... 6
  1.5.3 Atomistic Resolution ..................................... 7
  1.5.4 Coarse-Grained Resolution .............................. 7
  1.5.5 Standard MARTINI Model .............................. 8
CHAPTER 1. GENERAL INTRODUCTION

1.1 What is a Protein?

The activity within a living cell is based on a complex network of interacting biomolecules exchanging information and energy through biochemical processes. The amino acids are basic biomolecules that linked together by amide bonds between the amino group $-NH_2$ of one amino acid and the carboxyl group $-COOH$ of another form proteins. A protein can be thought of as a small apparatus dedicated to a specific task, e.g. membrane protein receptors provide mechanisms for information transfer between cells in complex signaling pathways. Proteins exhibit an extraordinary diversity of biomolecular roles ranging from enzymes, receptors, hormones, through protective, storage, or structural proteins to transport proteins. Classification of proteins is possible according to their function, and 3-dimensional (3D) structure or fold.

1.2 Protein Structure

Four different levels of structure are used when describing proteins. The first level is the primary structure, which is the amino acid sequence beginning at the N-terminus and ending at the C-terminus. In the majority of cases, the primary sequence completely determines how the protein will fold into its 3D structure. During the folding, specific secondary structure elements of the polypeptide backbone form regular patterns, such as $\alpha$-helices and $\beta$-strands. The spatial arrangement of the various secondary structure elements defines the tertiary structure which refers to an overall 3D shape. Several proteins can assemble together to yield a large multimer arrangement at the level of quaternary structure.

1.3 Malfunction of Proteins

Proteins need to be folded correctly in order to function optimally. Many diseases can be traced back to malfunctioning proteins. The cause of protein malfunction can be either intrinsic or extrinsic. The intrinsic malfunction results from a mechanism in the protein itself failing, causing rendering protein less effective or even inactive, and leading to diseases such as some types of cancer. The extrinsic malfunction is due to the failure of controlling mechanisms, i.e. proteins with a high propensity to misfold escape all the protective mechanisms and form aggregates within cells or (more commonly) in the extracellular space causing e.g. Alzheimer's and Parkinson's diseases that are directly related to the presence of aggregates of unfolded or misfolded proteins. Understanding the normal function and existing malfunction of proteins is of vital importance for current day medicine, and allows the search for therapies aimed at normalizing the functions of proteins. However, to use proteins as potential drug targets, their structure and mechanism of function should be known. The most common experimental methods for determining protein structure are presented in the following section.
1.4 Experimental Structure Determination

The first protein structure model of myoglobin was solved by X-ray crystallography in the 1958 by John Kendrew et al.\(^4\) To this day, 82,992 structures are deposited in the Protein Data Bank (PDB)\(^5\) solved by either the X-ray crystallography (88.0%), nuclear magnetic resonance (NMR) spectroscopy (11.5%), or molecular electron microscopy (EM) (0.5%). However, the EM delivers only low resolution structures, of large protein assemblies, while the X-ray crystallography and NMR solves structures mostly in high resolution, but of small size. The two latter methods are used on routine basis for determining protein structure, and thus will be described shortly in the two following subsections.

1.4.1 X-ray Crystallography

In the X-ray crystallography the 3D structure of a protein is obtained from a crystal that is exposed to an X-ray beam. The resulting individual diffraction patterns are recorded on an area detector during small rotational increments of the crystal and are combined into a set of diffraction patterns that can be processed, initially to yield information e.g. about the crystal packing symmetry. The intensities of the spots in the set of diffraction patterns are then used to determine the structure factors needed for calculating a 3D image of the electron clouds of the molecule called electron density map. Using a series of refinement techniques the quality of this map is improved until it is of sufficient clarity to build a molecular model using the protein sequence. The resulting structure is then refined to fit the map more accurately and a set of Cartesian coordinates for every non-hydrogen atom in the protein is derived.\(^6\) Although X-ray crystallography is a powerful method to elucidate the structure of a protein, the crystallization and refinement processes are nontrivial. The main problem for this method is the formation of a (meta)stable lattice of protein structures; proteins that possess unfolded tertiary structures with many degrees of freedom in solution cannot be crystallized. Even when a crystal can be obtained, the highly flexible or mobile regions, such as the N- and C-termini of the protein chain or flexible loops connecting secondary structure elements, are poorly defined or even absent in the electron density map, and can thus be modeled with only limited confidence. The quality of a crystal is also important, as protein crystals need to diffract well. The crystallization environment often happens to be different from the physiological environment resulting in production of structures that are not necessarily of the active forms of the proteins.\(^7\) In order to get insight to the function of proteins that is ultimately linked to their specific mechanic and chemical properties,\(^8,9\) we need to first know the structure.

\(^{3}\)www.pdb.org, July 11th 2012
1.4.2 Nuclear Magnetic Resonance

In NMR spectroscopy, protein structures are solved in solution, giving a more realistic view of the active protein. The NMR phenomenon is based on the existence of magnetic properties of atomic nuclei. When non-zero nuclei are placed in a magnetic field they align with the field and electromagnetic pulses can be used to probe the local chemical environment. For instance the geometric information obtained from spin-spin coupling arises from the interaction of different spin states of the nuclei through the chemical bonds, resulting in the splitting of NMR signals (J-coupling). The transfer of magnetization from one nuclear spin to another through dipole-dipole interactions (Nuclear Overhauser Effect - NOE) provides spatial information between two neighboring nuclei. The final, averaged model of a protein is developed by determining the mean position of each atom across an ensemble of models produced from analyzing the set of distances between specific pairs of hydrogen atoms that fits the NMR data. Finally, the average model is adjusted to obey normal bond distances and angles.

In comparison to X-ray crystallography, NMR contains a wealth of information about the dynamics of a protein in solution, thanks to information on the relaxation time of a given spin-state that depends on motions within the molecule. The different conformations are stored in a single NMR structure file. Historically, NMR was limited to structures of relatively small size. Even though new NMR techniques for studying proteins with a mass up to 100 kDa and DNA/RNA molecules up to 35 kDa are now possible, the size of the system is still much smaller than in the X-ray crystallography, which can handle complexes with a mass larger than 100 kDa (as long as suitable crystals can be obtained). Additionally, recent years have seen the emergence of solid-state NMR (ss-NMR), with which it is possible to use NMR for insoluble protein structures.

Each of the above experimental methods, has weaknesses, but both provide the high-resolution structures that are invaluable for probing protein dynamics and consequently the mechanism of action using computational modelling, notably molecular dynamics (MD) simulation techniques.

1.5 Computational Modelling

Computational methods can be divided into two main groups based on different hierarchical levels of accuracy. Quantum mechanics (QM) describes electrons explicitly, and is used to describe their movement in atoms (and molecules). QM is a very accurate method, however, both computational and time demanding, as well as limited to only small systems. On the contrary, in molecular mechanics (MM) the electrons are treated implicitly and each atom is treated as a single particle. Even though it allows the study of bigger systems including e.g. systems composed of several (macro)molecules, it is less accurate then QM. The work presented in this dissertation is performed using MM methods, and therefore QM methods will not be described any further.
1.5. COMPUTATIONAL MODELLING

1.5.1 Force Field

MM relies on three basic principles. The first, the thermodynamic hypothesis, states that a (macro)molecule driven by thermodynamic forces will change its conformation from the structure that represents a high energy state to a native structure which represents the global energy minimum state in a reversible fashion. The second principle, the additive assumption, states that the total potential energy, $V$, of a system can be determined as a sum of all the contributions from a number of separate potential energy functions with the spatial and geometrical coordinates as parameters. This collection of potential functions is called the force field, and is separated into terms representing bonded and non-bonded interactions between atoms. The bonded interactions are divided to terms involving two, three and four connected atoms. For example, the harmonic terms describing the distortions from equilibrium positions in bond-stretching and angle-bending can be calculated from:

$$V_{\text{bond}} = \sum_{i=1}^{N_b} \frac{1}{2} k_i (r_i - r_{0,i})^2 \quad (1.1)$$

$$V_{\text{angle}} = \sum_{i=1}^{N_\Theta} \frac{1}{2} k_\Theta (\Theta_i - \Theta_{0,i})^2. \quad (1.2)$$

Additionally, a term constructed from four connected atoms to describe the periodic torsional motion of dihedral angles can be calculated from:

$$V_{\text{dihedral}} = \sum_{i=1}^{N_\phi} \frac{1}{2} k_\phi \cos(n_i(\phi_i - \phi_{0,i}))^2 \quad (1.3)$$

The above three terms are sufficient to describe the bonded interactions. However, in some cases e.g. it is important to keep part of the molecule that involves an $sp^2$ carbon, planar. For these cases additional term is used known as the improper torsional term:

$$V_{\text{improper}} = \sum_{i=1}^{N_\xi} \frac{1}{2} k_\xi (\xi_i - \xi_{0,i})^2. \quad (1.4)$$

All of the above interactions are represented by harmonic potentials for the bond lengths $r_i$, bond angle $\Theta_i$, dihedral angle $\phi_i$, and improper dihedral angle $\xi_i$. The $k_b$, $k_\Theta$, $k_\phi$, $k_\xi$ denote the force constants for the bond-stretching, angle-bending, dihedral and improper dihedral angles terms, respectively.

The non-bonded interactions are considered between atoms that are more distant and are not connected by bonds. In most cases there are two types of non-bonded interaction: the short-range and long-range interactions. The short-range interactions are termed the van der Waals interactions and describe the repulsion of two atoms due to overlapping valence electrons and attraction due to induction and dispersion forces. These interactions are approximated in the 12-6 Lennard-Jones potential. The distance dependence of the repulsion term is proportional to $r_{ij}^{-12}$ inter-atomic distance mimicking the exponential
soft-wall behaviour, and proportional to \( r_{ij}^{-6} \) with regards to the attraction. The 12-6 Lennard-Jones potential on a given particle \( i \) due to particles \( j \) in a system is described by:

\[
V_{LJ}^{ij} = \sum_{i \neq j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right],
\]

(1.5)

where \( \epsilon_{ij} \) denotes the well-depth and \( \sigma_{ij} \) is a finite distance at which the inter-particle potential is zero.

The long-range interactions are a consequence of electric charges in the system. The individual atoms of a molecule can bear a charge, allowing for the use of Coulomb’s law to describe the mutual interactions between two (partial) atomic charges, and providing multipoles for molecules or individual charge groups. The Coulomb potential on particle \( i \) due to particles \( j \) reads:

\[
V_{\text{Coulomb}} = \sum_{i \neq j} \frac{1}{4\pi\epsilon_0\epsilon_{\text{rel}}} \frac{q_i q_j}{r_{ij}},
\]

(1.6)

where \( \epsilon_{\text{rel}} \) is dielectric constant of the medium. The electrostatic interactions decreases as \( r_{ij}^{-1} \) increases, making them longer ranged then the van der Waals interactions.

The third principle of MM, the transferability, is based on the assumption that parameters derive from small molecules such as bond length and angle, can be transfer to larger, more complex, macromolecular systems. Therefore, systems of different sizes can be studied within the same force field in a search for the lowest energy conformations, or to move molecules to sample the phase space, either randomly or deterministically by solving the equations of motions for each point mass.

### 1.5.2 Molecular Dynamics

The MD simulation method is often attributed to the pioneering work of Alder and Wainwright, who in late 1950s applied MD on hard spheres to investigate the properties of fluids. Since then, MD has been used to study different systems ranging in size and composition; from protein in water through proteins embedded in membranes composed of different lipids, surfactants, sterols, or detergents.

The basic principle behind MD is Newton’s 2nd law of motion used to calculate the dynamics of the system:

\[
\vec{F} = m\vec{a}.
\]

(1.7)

If the mass of each particle in the system is known and the forces are derived from the interactions with surrounding particles according to force field, the acceleration of each particle can be calculated. Then, the instantaneous velocity and the displacement can be calculated from:

\[
\frac{\vec{F}_i(t)}{m_i} = \vec{a}_i(t) = \frac{d\vec{v}_i(t)}{dt}
\]

(1.8)
and
\[ \vec{v}_i(t) = \frac{d\vec{r}_i(t)}{dt}. \] \hspace{1cm} (1.9)

The force acting on each atom \( i \) in the system is given by the negative gradient of the potential energy function \( V \), which depends on the coordinates of all other atoms in the system: \(^{14}\)
\[ \vec{F}_i(t) = -\frac{\partial V(\vec{r}_1(t), \vec{r}_2(t), ..., \vec{r}_N(t))}{\partial \vec{r}_i(t)}. \] \hspace{1cm} (1.10)

If the potential energy of the system is known and the coordinates for a starting structure and a set of velocities are given, then the force acting on each atom is calculated and a new set of coordinates is generated, from which new forces are calculated. Repetition of this procedure will generate a trajectory corresponding to the evolution of the system in time. \(^{14}\)

### 1.5.3 Atomistic Resolution

MD simulations with the atomistic resolution are well-established and deliver a generous amount of details and insights for the studied system allowing investigation of intra- and inter-molecular processes. In the atomistic description, each atom in the system is defined as a single interaction center, and the forces acting on it are evaluated every time step. The time step is limited by the fastest vibrations in the system (C-H), and thus corresponds to 1-2 fs. A time step of this size allows for stable numerical integration, however, limits the accessible time and length scales for biomolecular simulations. Thus, studies of conformational changes of large proteins are out of reach for atomistic simulations. Therefore, the use of full atomistic models is restricted to relatively small systems or short time scales. Often atomistic simulations are performed in combination with techniques to enhance sampling, such as replica exchange, \(^{23}\) meta dynamics, \(^{24}\) or with the use of implicit solvent model, which removes significant amounts of friction from the system, speeding up computations. \(^{25}\) The clear advantage of the atomistic simulation is that the results are detailed. The most widely used biomolecular force fields include the AMBER, \(^{26,27}\) CHARMM, \(^{28}\) GROMOS \(^{29}\) and OPLS \(^{30,31}\) force fields.

### 1.5.4 Coarse-Grained Resolution

The limitation of the relatively short time scale and size of the system of atomistic resolution can be overcome by removing degrees of freedom. One way of doing this is to group atoms and let them move together, removing the fastest vibrations in the system and allowing time steps of 10s of femtoseconds. This efficiently smooths the potential energy functions and accelerates the speed of the simulation. This approach is called coarse-graining (CG) and several different CG models for proteins have been developed during the years, \(^{32,33}\) ranging from models with one bead per amino acid \(^{34,35}\) to models with up to five beads per amino acid. \(^{36-38}\) The CG models are highly applicable for studying biological phenomena that are driven by hydrophobic and hydrophilic interactions \( e.g. \) protein mediated vesicle fusion, \(^{21}\) or membrane self-assembly around
membrane proteins. However, one has to be aware that the coarse description also ignores details which are central in keeping structural integrity of proteins and molecular interactions. The lack of directional hydrogen bonds in the CG model limits the study of detailed molecular interactions like protein-ligand binding (described in Chapter 2), and also requires the secondary structure constraints. Thus, protein structure has to be stabilized either by an application of harmonic restraints to mimic secondary structure H-bonds or by application of an elastic network model to avoid an unphysical collapse of the structure during the simulation.

The main CG model used in this dissertation is the MARTINI force field. Therefore the standard version of the MARTINI model will be described in more details in the next subsection.

1.5.5 Standard MARTINI Model

Among several CG models formulated for protein-lipid-water systems, the MARTINI force field has become very popular due to its large library of the biologically relevant building blocks, such as lipids, surfactants, sterols, organic solvents, polymers, sugars, and amino acids. This makes MARTINI model a good choice for simulating a broad range of different systems. The MARTINI model is based on an approximate four-to-one mapping. This means that on average four heavy atoms plus associated hydrogen atoms are represented by a single interaction center. The exception are ring-like molecules and small molecules (e.g. benzene or cholesterol) that are mapped with higher resolution - up to two-to-one. Each residue is mapped to one backbone bead (BB) placed at the center of mass (COM) of the backbone atoms: N, Cα, C, and O, and zero to four side chain beads (SC). There are four main types of beads: polar (P), non-polar (N), apolar (C), and charged (Q). The bead types can be further divided denoting their hydrogen-bonding capabilities: donor (d), acceptor (a), both, donor and acceptor (da), none (0). Furthermore their degree of polarity ranges from 1 for low polarity to 5 for high polarity. The ring beads are denoted, and thus SP* denotes the polar type of a ring bead, SN* the non-polar type of a ring bead, SC* the apolar type of a ring bead, and SQ* the charged type of a ring bead, where * is replaced by the hydrogen-bonding capabilities for N and Q bead types, or degree of polarity for P and C bead types. The mass of the beads is set to 72 amu, with the exception of the ring beads, for which the mass is set to 45 amu. The example mapping for alanine, lysine and tryptophan is shown in Figure 1.1.
The MARTINI force field has been developed in close connection with atomistic models, especially regarding the bonded interactions. The calibration of the non-bonded interactions of the chemical building blocks is done against experimental data, in particular thermodynamic data such as oil/water partitioning coefficients. Thus, the partitioning of amino acids between hydrophobic and hydrophilic environments is well described within this model. The bonded and non-bonded interactions are treated as described in Subsection 1.5.1. However, the strength of an interaction in the Lennard-Jones potential, Eq. (1.5), determined by the well-depth $\epsilon_{ij}$, depends on the type of interacting beads, and the effective size of beads $\sigma$, that is $\sigma = 0.47 \text{ nm}$ for normal types of particles and $\sigma = 0.43 \text{ nm}$ for ring-like molecules. In the Coulomb potential Eq. (1.6), which describes interactions between charged beads, a relative dielectric constant is set to $\epsilon_{rel} = 15$ for explicit screening. This is done as the CG water beads described by P4 bead type, which incorporate four water molecules, are electroneutral, and therefore do not possess any screening capabilities. To alleviate the implicit screening of electrostatic interactions, the polarizable water model was recently introduced for use with MARTINI. Here, water beads have two additional dummy particles bearing partial plus and minus charge, and thus the relative dielectric constant for explicit screening is lowered to $\epsilon_{rel} = 2.5$. The non-bonded interactions are cut off at 1.2 nm and smoothly shifted from 0.9 nm for the Lennard-Jones potential and from 0.0 nm for the electrostatic potential, to avoid noise in the simulations.

The CG procedure, as mentioned before, removes properties that make highly detailed interactions such as directional hydrogen bonding impossible, and thus hard to model in a straightforward manner. As this is crucial for simulating stable $\alpha$-helices and $\beta$-sheet structures, the secondary structure elements in MARTINI are stabilized using dihedral restraints on the backbone. This is sufficient for stabilizing the structure of peptides while for larger proteins, extra restraints on the structure need to be introduced. There are two approaches for doing this in the MARTINI model, both with the addition of
an elastic network model. In the first approach extra harmonic bonds between non-bonded beads based on a distance cut-off are added to the MARTINI topology file. The second approach, referred to as the ELNEDIN model,\textsuperscript{40} that constitutes some change to the MARTINI force field, uses a global elastic network between the backbone beads, which are within a specific distance in the initial structure. This model will be further discussed in Chapter 2.

This short introduction covers only the basics of the methods used in the following Chapters. The specific biological examples and methods for their study will be discussed in more detail within the specific project.
Chapter 2

Studying Protein Domain Movements - domELNEDIN Model

This chapter describes the domELNEDIN model, its development, and its application to a ligand-binding protein. The findings of this project have been compiled in a manuscript that has been submitted to Journal of Chemical Theory and Computation, see Appendix A.

Contents

2.1 Protein Domains .................................................. 12
2.2 domELNEDIN Model ............................................. 13
2.3 Test Example - Periplasmic Leucine-Binding Protein .......... 14
   2.3.1 LBP Domain Definitions ................................. 15
2.4 Methods .......................................................... 16
   2.4.1 LBP Setups ............................................... 17
   2.4.2 Simulation Details ....................................... 17
2.5 Protein Dynamics on the Nanosecond Time Scale ............... 17
   2.5.1 Stability of Protein Structure in the MARTINI Model ... 17
   2.5.2 ELNEDIN Scaffold ...................................... 18
   2.5.3 domELNEDIN Scaffold ................................. 24
2.6 Protein Dynamics on the Microsecond Time Scale .............. 25
   2.6.1 Starting from the Open Conformation .................. 25
   2.6.2 Starting from the Closed Conformation ................ 28
   2.6.3 Protein-Ligand Interactions ............................ 31
2.7 Conclusions ...................................................... 33
2.1 Protein Domains

In all large proteins, the polypeptide chain folds into several globular units, sometimes loosely connected, which are commonly referred to as domains or structural domains to indicate that they are identified from the 3D structure. Large movements of one protein domain relative to another provide spectacular examples of protein flexibility and are important for a variety of protein functions, like catalysis, regulation of activity, transport of metabolites, and formation of protein assemblies. Domains often close around a binding site shielding the substrate from water, surrounding it with catalytic residues, and preventing its escape. The presence of bound substrates stabilizes a closed conformation, and the absence of substrate favors an open conformation.

There are two basic ways for domains to move (Figure 2.1). First the shear motion involves small 1 Å to 2 Å sliding movements between close-packed segments of polypeptide, e.g. helices, across and parallel to the plane of their interfaces. The main chain atoms involved in shear motions are constrained by closed packing, while the interface side chains usually have torsion angle changes less than 20° and retain the same packing configuration throughout the motion. Thus, the shear motion is the result of many small local motions, and has been recognized in citrate synthase, hexokinase or alcohol dehydrogenase. The second type of motion called the hinge motion, occurs in strands, β-sheets, and α-helices that are not constrained by tertiary packing interactions. In contrast to the shear motion, the most basic motion of a polypeptide chain is a few large changes in main chain torsion angles in a localized region, i.e. at the hinge, consequently leading to rotation of polypeptide chain by up to 60°. The classical examples of hinge motion are recognized in periplasmic binding proteins, adenylate kinase, or in lysozyme.

![Figure 2.1](image-url)
2.2 domELNEDIN Model

The inter-domain interactions are expected to be driven by hydrophobic and hydrophilic interactions rather than hydrogen bonding. Therefore, the MARTINI model\textsuperscript{37,38} should be suitable for studying the conformational changes associated with domain movements. However, as mentioned in Chapter 1, Subsection 1.5.5, the coarse description of the system forces the use of a global elastic network model - the ELNEDIN model,\textsuperscript{40} to keep the tertiary structure of especially large proteins stable. This model constitutes changes to the MARTINI force field, \textit{e.g.} the backbone bead of a residue in the ELNEDIN model is placed at the location of the \(C_\alpha\) atom rather than at the location of the COM of the backbone atoms (N, \(C_\alpha\), C, O), as it is done in the MARTINI model. The new position of the backbone bead has led to the need for re-parametrizing the bonded parameters (bond and angle equilibrium values and the corresponding force constants) for all beads. Additionally, also the representation of the ring structures of amino acids has been modified. For both phenylalanine and tyrosine an extra bond was used to maintain the ring structure, and in the case of histidine and tryptophan the asymmetry in their rings was considered.\textsuperscript{40}

In the ELNEDIN model an elastic network is applied to the whole protein with bonds connecting all CG backbone beads, which are within a specific distance in the initial structure. Two parameters are used to define the ELNEDIN network and to tune the dynamics. These parameters are: the cut-off distance between the backbone beads \(R_C\) (nm), which describes the range of beads that can be connected with the additional elastic bonds, and the spring force constant \(K_S\) (kJ⋅mol\(^{-1}\)⋅nm\(^{-2}\)), which describes the stiffness of the elastic bonds. The default parameters for the ELNEDIN model, suggested to reproduce protein dynamics comparable to atomistic simulations are \(R_C = 0.9\) nm and \(K_S = 500\) kJ⋅mol\(^{-1}\)⋅nm\(^{-2}\).\textsuperscript{40} However, using the ELNEDIN model is not optimal, as the global elastic network imposes a strong bias towards the initial structure,\textsuperscript{40} thereby greatly limiting the description of the protein dynamics otherwise accessible on the microsecond time scale. For this reason, we proposed an alternative extension to MARTINI-ELNEDIN model, referred to as the domELNEDIN model.

In the domELNEDIN model protein domains are allowed to move independently, and thus conformational changes in an unbiased manner during an MD simulation are allowed. The elastic network is applied to backbone beads only locally in the protein domains (Figure 2.2) and, as in the case of the ELNEDIN model, the elastic network is set up using the cut-off distance between the backbone beads \(R_C\) (nm) and the spring force constant \(K_S\) (kJ⋅mol\(^{-1}\)⋅nm\(^{-2}\)). The domELNEDIN model uses the ELNEDIN mapping in combination with the MARTINI force field.

It is expected that applying an elastic network model locally in the protein domains is sufficient for stabilizing the protein structure, to both avoid collapse of the structure and to keep it in its native conformation. The parameters for setting up an elastic network for the domELNEDIN model are tested and discussed in the following sections.
CHAPTER 2. STUDYING PROTEIN DOMAIN MOVEMENTS - DOMELNEDIN MODEL

Figure 2.2 - Comparison of elastic network models generated with $R_G = 0.8$ nm for the ELNEDIN and domELNEDIN models shown on the example of 1USK (test example, described in Section 2.3). The protein structure is shown as a gray Licorice representation of backbone beads. Bonds are colored as follows: gray - bonds present in both models, red - bonds not included in the domELNEDIN model.

2.3 Test Example - Periplasmic Leucine-Binding Protein

The periplasmic leucine-binding protein (LBP) has been chosen as a test case for the domELNEDIN model, as a classical example of protein exhibiting a hinge motion, which has been referred to as the Venus flytrap. LBP belongs to the superfamily of binding proteins. More than 100 crystal structures of periplasmic binding proteins (PBP) from different sources and in different conformations have been solved, showing a conserved structural fold, despite high sequence diversity. The protein fold consists of two domains connected by a three-, two- or one-stranded hinge (group I, II and III of PBP, respectively). LBP is a group I PBP, and LBP structures have been established for both open conformations and closed conformation with the leucine ligand bound in the cleft between the domains. In gram-negative bacteria, binding proteins found in the periplasm, such as LBP, act as primary receptors in adenosine triphosphate (ATP)-binding cassette (ABC) transport systems, trafficking various substrates across the plasma membrane at the expense of ATP hydrolysis. Despite the great variety in substrates ranging from nutrients such as sugars and amino acids in prokaryotes to poly-saccharides, lipids and hormones in eukaryotes, the ABC transport systems seem to share overall mechanism. For the ABC-importers present in gram-negative bacteria such as E.coli, a binding protein entraps the substrate in a closed conformation, and docks in this form to the transmembrane domain (TMD) which is interacting closely with the ATP binding domains (ATP_BD) on the cytoplasmic site. The ATP-binding and hydrolysis then help fuel the conformational shifts necessary to open the binding protein, release the substrate to the TMD, change the TMD to an occluded state, and finally release the substrate to the cytoplasm (see Figure 2.3).
2.3. TEST EXAMPLE - PERIPLASMIC LEUCINE-BINDING PROTEIN

The role of the substrate in the mechanism for the different ABC transporters, as well as the complex interplay between conformational changes and ATP interaction have still not been settled on.\textsuperscript{57-60}

### 2.3.1 LBP Domain Definitions

The division of a protein into structural domains will often be done in the most sensible way by a trained eye,\textsuperscript{60} and the well established fold databases CATH\textsuperscript{61} and SCOP\textsuperscript{62} rely on human expertise. However, with the pace at which new protein structures are submitted to the protein data bank (PDB),\textsuperscript{5,63} the human expert assignments lag behind, and the development of methods for high-quality automatic assignment of domain boundaries is an active field of research.\textsuperscript{60}

Four different domain definitions for LBP were therefore tested in this study (see Table 2.1). Based on visual inspection of both the open and closed conformations of LBP (PDB: 1USK\textsuperscript{51} and PDB:1USG,\textsuperscript{51} respectively) it was decided to define domain 1 as residues 1-120 and 250-330 and domain 2 as residues 121-249 and 331-345. This is referred to as the “main” domain assignment, and is identical to the assignment by CATH. Apart from this, “main loose” definition, where four amino acids in each of the three domain linkers were released from the elastic bonds was applied. Two alternative pro-
tein domain boundaries were also acquired from two public servers for automatic protein
domain assignment. One is the pDomains server,\(^{64,65}\) which for a protein PDB Id can
provide and overview of domain assignment results from seven different methods, as well
as provide a consensus assignment based on these results. In the consensus assignment,
the different methods contribute with a weight which is automatically adjusted based on
benchmarked knowledge of each method's performance for that particular type and size of
protein.\(^{60,66,67}\) For the closed conformation two consensus assignments were presented,
the first consensus assigned it as a one-domain protein while the second consensus was
a two-domain assignment. The two-domain consensus assignments for the open and
closed structures used in this study were identical, except for residue 333, which was as-
signed to domain 1 in the closed conformation and to domain 2 in the open conformation.
As seen in Table 2.1, we used the assignment as given to the open conformation. The
other server for automatic protein domain assignment is the DomFOLDpdp server,\(^{68,69}\)
which based on the amino acid sequence predicts the protein fold using the nFOLD3
method,\(^{70}\) and then analyze the best fold model with the Protein Domain Parser (PDP)
program.\(^{71}\)

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>LBP Domain Boundary Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1</td>
<td>Domain 2</td>
</tr>
<tr>
<td>&quot;main&quot;</td>
<td>1-120, 250-330, 121-249, 331-345</td>
</tr>
<tr>
<td>&quot;main loose&quot;</td>
<td>1-118, 252-328, 123-247, 333-345</td>
</tr>
<tr>
<td>pDomains(^{64,65})</td>
<td>1-124, 247-332, 125-246, 333-345</td>
</tr>
<tr>
<td>DomFOLDpdp(^{68,69})</td>
<td>1-119, 249-332, 120-248, 333-345</td>
</tr>
</tbody>
</table>

In the present study we examine the ability to produce a stable structural scaffold for
both the open and closed conformation of LBP on the 100 ns time scale for the standard
MARTINI, the ELNEDIN and the domELNEDIN models, by comparing the structure
and dynamics to atomistic simulations. Following this, microsecond long domELNEDIN
simulations are carried out for both the open and closed conformations, to study the
conformational flexibility of LBP in water. To illustrate the differences between the
ELNEDIN and domELNEDIN models on the long time scale, results from the corre-
spanding ELNEDIN simulations are also reported. Additionally, to inspect the sensitivity
of the domELNEDIN model towards the assigned protein domain boundaries, four
different domain setups are tested. Finally, the model limitations as well as the biological
implications of the results are discussed.

2.4 Methods

All details on the setup and simulations can be found in the Method section of the
manuscript in Appendix A. Additionally, the protocol for setting up domELNEDIN
simulations can be found in Appendix B. Here only a brief outline will be provided.
2.5. PROTEIN DYNAMICS ON THE NANOSECOND TIME SCALE

2.4.1 LBP Setups

MD simulations of LBP were carried out starting from closed (PDB: 1USK\textsuperscript{51}) or open (PDB: 1USG\textsuperscript{51}) conformation of the protein. For both conformations two different setups were applied; either with or without the leucine ligand present in the binding cleft.

2.4.2 Simulation Details

The GROMACS package version 4.0.\textsuperscript{72,73} was used for all simulations. The atomistic simulations were performed using the AMBER03 force field,\textsuperscript{74} and the CG simulations using the MARTINI force field version 2.1,\textsuperscript{38} in combination with the ELNEDIN\textsuperscript{40} or domELNEDIN model.

Each structure, both atomistic and CG, were placed in a cubic box with dimensions of 100 Å. Counter ions were added (9 Na\textsuperscript{+}) and the protein was solvated using the SPC water model\textsuperscript{75} for the atomistic simulations and water beads\textsuperscript{42} for the CG simulations. Systems were energy minimized before a production run of 100 ns with a 2 fs time step for the atomistic simulations. For the CG models the time steps were as follows: 25 fs for the MARTINI model, and 10 fs for the ELENDIN and domELNEDIN models. These were run for 25 ns or 1 µs, depending on the purpose of the simulation, which correspond to 100 ns and 4 µs of “real” time.\textsuperscript{37,70} The pressure and temperature were kept constant at 1 bar and 300 K, respectively, for all simulations, using the Berendsen coupling algorithm.\textsuperscript{77}

The root mean-square deviation (RMSD), the RMSD of individual residues, the root mean-square fluctuations (RMSF) of individual residues, and principal component analysis (PCA),\textsuperscript{78} were calculated for the protein backbone based on the \( \text{C}_\alpha \) atoms for the atomistic simulations or the backbone beads for the CG simulations.

2.5 Protein Dynamics on the Nanosecond Time Scale

2.5.1 Stability of Protein Structure in the MARTINI Model

In the MARTINI CG model, even though detailed interactions are lost, the resolution is high enough to represent the particular trends of the different amino acids, due to the different bead types used to describe them. Additionally, because of the calibration of the non-bonded interactions against experimental data, the partitioning of amino acids between hydrophobic and hydrophilic environments is well described within this model.\textsuperscript{38} However, as the CG description removes the fastest vibrations in the system, the secondary structure elements are not self-contained, lacking the directional hydrogen bonds, and thus the helices or extended structures are maintained using backbone angle and dihedral restraints.\textsuperscript{38} This is sufficient for purely \( \alpha \)-helical structures, as saposin like proteins\textsuperscript{79-82} (described in Chapter 5) or membrane proteins completely surrounded by a lipid membrane.\textsuperscript{83-86} However, when the protein structure contains \( \beta \)-sheet elements, the MARTINI model has a hard time maintaining the tertiary structure of the protein,
as the $\beta$-sheet interactions are non-local in nature and therefore, poorly described by this model.

To see if this is also the case for the LBP protein, which is a globular protein, with each of its two domains consisting of a central $\beta$-sheet flanked by $\alpha$-helices, we set up 100 ns MARTINI simulations for both open and closed conformations of LBP. To no surprise, the MARTINI simulations for both the open and closed crystal structures showed the tertiary structure collapse into tightly packed globular structure. To exclude that the protein structure resulting from the MARTINI simulations was in fact the closed conformation of the protein, the final structure was compared to a known closed crystal structure of LBP, however, with a difference in RMSD value of 5.6 Å it did not resemble the known closed form, confirming tertiary structure collapse.

The inability of the standard MARTINI model to keep the structure stable for even 100 ns makes this model irrelevant for further studies of LBP, and shows the necessity of application of an elastic network model to LBP. Thus, in the next sections, the focus will be put on the ELNEDIN and domELNEDIN models.

### 2.5.2 ELNEDIN Scaffold

In the ELNEDIN model the elastic network is applied to the backbone stabilizing the overall protein structure, allowing for structural fluctuations on the nanosecond time scale, comparable to those observed for atomistic simulations. The elastic network is controlled by adjusting the cut-off distance $R_C$ for generating bonds and the force constant for stiffness of the elastic bonds $K_S$. The default values for the ELNEDIN model are $R_C = 0.9$ nm and $K_S = 500$ kJ·mol$^{-1}$·nm$^{-2}$, however, it was suggested that the elastic network model for each protein can be different. Thus, for the purpose of establishing the optimal set of parameters to describe the dynamics of LBP, nine different simulations were carried out varying the cut-off distance $R_C$ (nm) ∈ \{0.8, 0.9, 1.0\} and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) ∈ \{50, 500, 5000\}, starting from the crystal structure of open unliganded conformation of LBP (PDB: 1USG$^{51}$) and closed liganded conformation of LBP (PDB: 1USK$^{51}$). The RMSD, and RMSF per residue showed that the protein dynamics for both, open and closed crystal structures, decrease with the increase of the cut-off and force constant values, as expected (Figure 2.4 and Figure 2.5). The best overlap between the protein dynamics observed in the atomistic and the ELNEDIN simulations of LBP is given for a set of parameters $K_S = 500$ kJ·mol$^{-1}$·nm$^{-2}$ and a cut-off of $R_C = 0.8$ nm. However, for further tests both cut-off values will be used, the $R_C = 0.8$ nm and the default cut-off for the ELNEDIN model $R_C = 0.9$ nm.
2.5. PROTEIN DYNAMICS ON THE NANOSECOND TIME SCALE

Figure 2.4 - The (A) RMSD as a function of simulation time and (B) RMSF as a function of residue number for ELNEDIN simulations of the closed conformation of LBP with ligand (red line). Nine different elastic networks were tested, varying the cut-off distance $R_C$ (nm) ∈ {0.8, 0.9, 1.0} and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) ∈ {50, 500, 5000}. Plots include results from the atomistic simulation (black line) for comparison of protein dynamics.
Figure 2.5 – The (A) RMSD as a function of simulation time and (B) RMSF as a function of residue number for ELNEDIN simulations of the open conformation of LBP without ligand (red line). Nine different elastic networks were tested, varying the cut-off distance $R_C$ (nm) $\in \{0.8, 0.9, 1.0\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\in \{50, 500, 5000\}$. Plots include results from the atomistic simulation (black line) for comparison of protein dynamics.
Choice of Starting Structure

When applying the ELNEDIN model it is important that the initial structure represents a stable conformation, as the bonds forming the elastic network are put on the initial structure, and the original distances are used as the equilibrium distances for the elastic bonds. In other words the conformation will be biased towards the original structure, and no major structural changes can take place. The choice is then either to use the crystal structure of the protein or to carry out a short equilibration in atomistic resolution to obtain an equilibrated structure before imposing the elastic network bonds. In this way the ELNEDIN scaffold will vary in the number of existing bonds, as during the atomistic simulation, the distances will change following the definition of used force field, in this case will follow AMBER03 parameters. This means that the equilibrium distances for the elastic bonds that were formed based on crystal structure below cut-off may now change the value and be above the predefined cut-off, or vice-versa.

![Figure 2.6](image.png)

**Figure 2.6** – The running average of RMSD for the protein backbone in the atomistic simulation of closed (black line) and open (red line) conformations of LBP.

The structures in the atomistic simulations were stable reaching a plateau for the backbone RMSD of both the open and the closed conformation of LBP already after 9 ns of simulation (Figure 2.6). To examine the sensitivity of the sampled dynamics to the choice of initial structure, simulations were carried out where the ELNEDIN model was either based on the crystal structure (X-ray) or the equilibrium structure averaged over 40 frames spanning the time between 9 and 11 ns of atomistic simulation (AAeq). The RMSD and the RMSF for individual residue (Figure 2.7) show that the overall change of the structures observed during the ELNEDIN simulations is on the same level no matter which starting structure was chosen. Only in the RMSD per residue plot for the open conformation of LBP the difference depending on the initial structure is significant, showing that a better fit with atomistic simulations is given when the crystal structure is used. When looking at the PCA plots (Figure 2.8) the biggest difference in flexibility is given when using the crystal structure with the elastic network scaffold generated for cut-off $R_C = 0.8$ nm, compared to the simulations starting from the same structure, however, using cut-off $R_C = 0.9$ nm, for both the open and closed conforma-
CHAPTER 2. STUDYING PROTEIN DOMAIN MOVEMENTS - DOMELNEDIN

MODEL

Figure 2.7 - RMSD as a function of simulation time and RMSF and RMSD per residue obtained from ELNEDIN simulations of the closed conformation of LBP with ligand (A) and the open conformation without ligand (B). The force constant parameter, $K_S$, of the elastic network is in all cases 500 kJ·mol$^{-1}$·nm$^{-2}$ and the cut-off, $R_C$, is either 0.8 nm or 0.9 nm as indicated on the figure. Plots include results for simulations starting from either the crystal structure (X-ray, red line) or the average structure, as explained in the text, (AAeq, orange line).

Figure 2.8 - The PCA analysis showing projection on the two first eigenvectors for AA simulations and ELNEDIN simulations. For ELNEDIN simulations results are shown for both structures, the crystal structures and equilibrated structure, for which the elastic scaffold was generated using $R_C = 0.8$ nm (black circles) and $R_C = 0.9$ nm (red circles) cut-off values.
tions of LBP. The explanation of this can lie in the number of generated elastic bonds. Firstly, the elastic scaffold built with cut-off $R_C = 0.8$ nm is less dense ($\approx 1100$ elastic bonds) than the one built with cut-off $R_C = 0.9$ nm ($\approx 1700$ elastic bonds), allowing for more flexibility in the proteins. Secondly, starting from the X-ray structure the difference in the number of elastic bonds, using $R_C = 0.8$ nm or $R_C = 0.9$ nm cut-off values is 69 and 51 for open and closed conformations of LBP, respectively. When looking at the elastic network scaffold built on the equilibrated structures the corresponding values decrease to 24 and 45 for open and closed conformations of LBP, respectively, resulting in less pronounced difference in the conformational space for this structures (Figure 2.8).

The placing of the elastic bonds also seems to influence the dynamical properties of the protein. Figure 2.9 shows the comparison of the elastic network model generated for the crystal and equilibrated structures of the closed and open conformations of LBP, using $R_C = 0.8$ nm. For the closed structure more elastic bonds specific for the X-ray structure (99) are placed between the domains, while the specific elastic bonds for the equilibrated structure (72) are rather placed inside the domains. When looking at the open conformation the situation is not as obvious, however, it is clear that the elastic bonds generated only for the X-ray structure (85) are stabilizing the hinge region, while the bonds generated only for the equilibrated structure (76) lock the domain to the hinge.

![Figure 2.9](image)

The sensitivity towards the choice of initial structure that is presented here for the ELNEDIN model is troubling, however, it is clear that for LBP, simply starting from the crystal structure compares better to the nanosecond dynamics in the atomistic simulation than starting from the equilibrated structure in the AMBER03 force field. Despite quite different overall levels of dynamics for the open and closed conformations of LBP, the ELNEDIN setup succeeds in reproducing a structural scaffold similar to what is seen in atomistic simulations of 100 ns, with $R_C = 0.8$ nm and $K_S = 500$ kJ · mol$^{-1}$ · nm$^{-2}$ for both LBP conformations. However, it does not allow studies that involve rearrangements of the protein domains, biasing the structure towards the initial conformation.
2.5.3 domELNEDIN Scaffold

The structural difference between different protein conformations is seen mainly to arise from the hinge, shear, or rotational motions between these structural domains. The RMSD between the crystal structures of the open and closed conformations of LBP used for this work is 7.04 Å, while the RMSD values between the same domains of these two conformations are 0.73 Å for domain 1 and 0.63 Å for domain 2, when using the “main” domain assignment (Table 2.1). The structural rearrangements between the conformations are thus mainly caused by movements in the hinge region of the protein. For other globular proteins, where we have compared crystal structures for different conformations, it is also clear that structural domains can be assigned which show a stable core no matter the conformation. For LBP and many other multi-domain proteins, the individual protein domains could therefore be restrained with elastic network bonds as in an ELNEDIN model, while still allowing for a full description of the structural ensemble covering the whole functional cycle. This is the concept of the domELNEDIN model; an elastic network is applied within each domain with the same type of parameter setup as for ELNEDIN, the difference being that no elastic network bonds are applied connecting the protein domains. For LBP the differences between ELNEDIN and domELNEDIN are seen in Figure 2.10 for both the open and the closed conformations.

The domELNEDIN elastic scaffold was built using $R_C = 0.8$ nm or $R_C = 0.9$ nm and $K_S = 500$ kJ·mol$^{-1}$·nm$^{-2}$. Since the protein domains are now allowed to move independently, it is not expected that the dynamics will be dampened in a systematic manner.
manner, as observed for the ELNEDIN model. Results indicate that the best fit with atomistic simulations is achieved when applying elastic scaffold parametrized with $R_C = 0.8$ nm, as in the case of ELNEDIN model. When the RMSD and RMSF values are compared for these two models (Figure 2.10), it is seen that the protein structure in the domELNEDIN simulations shows more flexibility. However, when comparing the RMSD and RMSF per residue between the ELNEDIN and domELNEDIN models, it is clear that the structural stability of the protein is on the same level on the 100 ns time scale.

On the nanosecond time scale the structural integrity and dynamics of the protein in the domELNEDIN model is equivalent to that observed for the ELNEDIN and atomistic simulations, using an elastic network scaffold parametrized with the $R_C = 0.8$ nm and $K_S = 500 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ for both LBP forms. The flexibility that domELNEDIN introduces to the protein scaffold should allow for the simulations of conformational changes on the microsecond time scale.

2.6 Protein Dynamics on the Microsecond Time Scale

For a number of the binding proteins it has been established that the protein structure without substrate occupies a wide range of conformations ranging from a full opened to a closed conformation. A closed conformation without substrate has not yet been observed for LBP, and to study the conformational flexibility of LBP, microsecond long domELNEDIN simulations have been carried out starting from both the open and closed unliganded conformations. For both cases, the domELNEDIN simulations are compared to the corresponding ELNEDIN simulations and the sensitivity towards the domain boundary definitions and the differences in the structure bias related to differences in equilibrium angles and elastic network bonds are tested.

2.6.1 Starting from the Open Conformation

The ELNEDIN model applied to the open conformation was expected to result in a stable structure. As there are only a few elastic bonds connecting the domains, moreover positioned in the hinge region, the tertiary structure change is allowed, however, limited, as further change towards the closed conformation will be energetically highly unfavorable due to the necessary change in the elastic bonds between the domains. The change of the open conformation towards the closed one goes from an RMSD of 7.0 Å to 5-6 Å with respect to the crystal structure of the closed conformation. On the contrary, for the domELNEDIN simulations of the open conformation the structure is free to close up. The closing of the open conformation, clearly approaches the closed conformation to a significantly higher degree than seen for the ELNEDIN simulations, with an RMSD as low as 3 Å with respect to the crystal structure of the closed conformation (Figure 2.11). Even though the LBP domains are structurally very similar between the conformations, there are a few differences in their modelling parameters since equilibrium
angle values for the backbone as well as the elastic network bonds are assigned based on the exact atom positions in the initial structures. Out of the ≈1080 bonds forming the elastic network for the domELNEDIN simulations ≈50 are unique for the closed or open structures (around 10 of these involve residues in the domain linkers) while the rest are shared between the setups for the two conformations. For eight of the backbone bend angles the equilibrium values differ by more than 10 degrees, and four of these are found in the linker region.

A setup was made named “topology swap”, where the topology input for the closed conformation was applied to the open conformation, enforcing the elastic network bonds and equilibrium angles of the closed conformation. For this setup, a simulation starting from the open conformation is seen to reach a closed structure which has an RMSD as low as 1.6 Å compared to the crystal structure of the closed form (Figure 2.11). The application of the structural parameters from the known closed structure thus allowed the proper “induced fit” of the domain structures into the closed conformation of the protein.

![Figure 2.11](image)

**Figure 2.11** – Backbone RMSD in 4 µs domELNEDIN simulations of the open conformation of LBP without ligand. Results are shown for two variations of the setup, namely the original with “main” domain assignment (red), and a setup where the topology of the closed conformation is applied in the simulation of the open conformation (black). The RMSDs as compared to the crystal structures of both the open and closed conformations are given.

To inspect if the outcome of the domELNEDIN simulations of LBP depends on the domain boundary assignment, besides the “main” domain assignment three different assignments have been tested (Table 2.1). The “main loose” assignment, as well as two domain assignments from the DomFOLDpdp server\(^68,69\) and the pDomains server\(^64,65\) were used. Based on these single simulations it is seen (Figure 2.12) that both proteins with domELNEDIN scaffold built based on the pDomains and DomFOLDpdp domain assignments close up with the RMSD value of ≈4 Å compared to closed conformation of LBP, while protein with domELNEDIN scaffold built with “main loose” domain assignment closes up to RMSD value of ≈3 Å compared to closed conformation of LBP similarly to protein with “main” domain assignment (Figure 2.11).
2.6. PROTEIN DYNAMICS ON THE MICROSECOND TIME SCALE

Figure 2.12 – Backbone RMSD in 4 µs domELNEDIN simulations of the open conformation of LBP without ligand. Results are shown for three variations of the setup, the “main loose” (red), the pDomains (orange), and the DomFOLDpdp (black) domain assignments. The RMSDs as compared to the crystal structures of both the open and closed conformations are given.

For several of the binding proteins, it has been established that the ligand free structure is very flexible, with a structural diversity going all the way from a completely open to a closed conformation, and with an estimated open-to-closed dynamics on the nanosecond to microsecond time scales. This is also what we observe for LBP. Both atomistic and domELNEDIN simulations, starting from the open conformation, show a highly dynamical structure (Figure 2.6 and Figure 2.14), yet fluctuating around a fully open conformation on the 100 ns time scale. In the domELNEDIN simulations, extended to the microsecond time scale, conformational changes towards a closed conformation are clearly observed. For the binding proteins where structural information on a closed conformation without ligand has been established, cases where the ligand bound and unbound structures are almost identical are observed, as well as cases where the closed conformation without substrate seem to close up in a distinguishable different conformation than when the substrate is present. In the domELNEDIN simulations, the closed structure obtained from the simulations starting from the open conformation is still around 3-4 Å different, depending on the domain assignment, from the crystal structure of the closed structure with ligand bound. That the RMSD does not go below 3 Å can either be because LBP actually closes up in a different manner when ligand is not present, or due to limitations in the coarse domELNEDIN model. To examine if the observed open-to-closed transition was dependent on the exact domain definitions, four different definitions were tested, including a setup where the linker region was released from the elastic network bonds. In all four simulations the open conformation changes to a closed structure similar to each other, however, these structures are still 3-4 Å different from the closed crystal structure of LBP. To elucidate if the different domain definitions would alter the frequency of which the structure changes from open to closed, a more extensive sampling is needed.

While the closing up of the open ligand free structure seems independent of the exact domain definitions, another limitation of the domELNEDIN model is that the elastic network bonds set up inside the protein domains are based on the original structure, and any “induced fit” going from an open to a closed structure will thus not be supported.
As the RMSD values between the domains in the closed and open structures are 0.73 Å for domain 1 and 0.63 Å for domain 2, this effect was not expected to be significant. However, when the equilibrium distances between backbone beads inside the domains corresponding to the closed conformation, were applied to the open conformation (“topology swap”), LBP was observed to undergo a conformational change all the way from the open conformation to a structure with RMSD of only 1.6 Å with respect to the crystal structure of the closed conformation. This shows that “induced fit” of the domains indeed play a role in the complete closure of the protein, but from these simulations it cannot be determined if this fit is induced by interaction with the substrate or if it could also take place in the ligand free situation.

2.6.2 Starting from the Closed Conformation

In all simulations of the closed conformation of LBP, LBP stays more or less closed. In the ELNEDIN simulations it is under no circumstances expected to observe any opening of the protein, as several elastic network bonds are applied between the domains, keeping them tightly together (Figure 2.9). Also for the domELNEDIN simulations the closed conformation is very stable and no opening is observed. Thus, even though the protein is free to change its conformation in a long time scale simulation, it stayed in the stable closed conformation as also observed in the atomistic simulation.

In the simulation starting from the closed conformation, but applying the structural parameters from the open form (“topology swap”), the protein clearly remains closed while re-modelling the internal domain structure, thus resulting in an elevated RMSD to the closed structure while keeping an even higher RMSD to the open structure (Figure 2.13). This observation corroborates that the application of the topology from a different conformation does not force a conformational change in itself, but merely introduce the “induced fit” of the internal domain structure corresponding to the opposite conformation.

**Figure 2.13** - Backbone RMS in 4 µs domELNEDIN simulations of the closed conformation of LBP without ligand. Results are shown for two variations of the setup, namely the original with “main” domain assignment (red), and a setup where the topology of the open conformation is applied in the simulation of the closed conformation (black). The RMSDs as compared to the crystal structures of both the open and closed conformations are given.
The same general trends are observed for the simulation with different domain assignments (Figure 2.14). The release of the linkers from the elastic network bonds of the protein domains in the “main loose” setup does not affect the overall structural changes, as well as none of the two other setups, namely pDomains and DomFOLDpdp (Figure 2.14), and the domELNEDIN model thus supports a stable structure even when additional freedom to change the structure is introduced.

Figure 2.14 - Backbone RMSD in 4 µs domELNEDIN simulations of the closed conformation of LBP without ligand. Results are shown for three variations of the setup, the “main loose” (red), the pDomains (orange), and the DomFOLDpdp (black) domain assignments. The RMSDs as compared to the crystal structures of both the open and closed conformations are given.

The simulations starting from the closed conformation with the leucine ligand removed from the setup shows a very stable structure, also on the microsecond time scale (Figure 2.13 and Figure 2.14). Comparing the atomistic simulations of the closed conformation with and without the leucine ligand present in the binding cleft indicates that the structure is even slightly more stable when the ligand is not present (data not shown). However, it cannot be ruled out that the observation of the microsecond stability of the closed conformation is due to an over-stabilization of protein-protein interaction in the MARTINI force field. Previous studies showed that in the standard MARTINI model the charged residues, as well as phenylalanine, proline, and to a lesser extent tryptophan, as well as the aromatic side chains, are found to be too hydrophobic. Additionally, it was also confirmed by studies of the free energy of amino acid side chains dimerization in different solvents with different polarity that the charged and polar interactions in a low dielectric medium are underestimated compared to the atomistic models. In a very recent update to the MARTINI protein modelling (version 2.2P), the description of side chains are further improved by adding particles with charges on the polar side chains as well as moving the charge on the charged side chains away from the van der Waals centre of the charged bead, to allow a better modelling for charge and polar interactions. The water model is also extended with particles with opposite and movable charges, to allow for a polarization of the water bead. Using this new MARTINI protein and water model together with domELNEDIN on the closed conformation with the ligand removed, microsecond long simulations still show a stable structure (see Chapter 3). This can arise from the fact that the improvements to MARTINI model only concern the side chains and not the backbone, the secondary and tertiary protein
structure is still stabilized by restraints. Furthermore, the backbone bonded parameters are dependent on predefined secondary structures of proteins fixed at strand, helix, or coil structures throughout simulations through the use of angle and dihedral potential energy functions imposed during the conversion from atomistic to CG representation. While the ability of the backbone of the protein to be flexible is often required during simulations since biological processes, such as protein folding and aggregation, typically involve transitions between different types of secondary structure. The simulations starting from the closed ligand free structure are artificial in the sense that the structure used for the setup contained the ligand. It cannot be ruled out that the protein would actually never close up exactly in this way if leucine was not present; this structure is not yet known. From the few crystal structures of full ABC-transporter complexes, where the interaction between a transmembrane permease and its binding protein can be seen, the two subunits of the permease interact with each their domain of the binding protein, making it plausible that the closed binding protein docks to the permease in its inward facing conformation. Then, an ATP-driven conformational change from the inward to an outward facing form of the permease breaks open the binding protein, for the substrate to be released into the permease, and by another ATP-driven conformational change release the binding protein in its open form into the solvent again (Figure 2.3). The permease interaction will by this mechanism help the binding protein overcome the energy barrier going from the closed to the open conformation. Even if the closed ligand free structure of LBP, obtained from the domELNEDIN simulations, does not exactly match the closed crystal structure with ligand bound, as we see in the domELNEDIN simulations, this leaves the question how the permease can distinguish between the substrate bound and empty closed conformations. However, modelling the kinetics of a binding-protein dependent transport system it is clear that the substrate-bound and empty binding proteins compete for the permease interaction. It has been shown also for the histidine-binding protein that ligand bound and ligand free proteins have equal affinity for the membrane-bound complex and that ligand free binding protein competes efficiently with ligand bound protein, and thus inhibits the substrate transport.

On the microsecond time scale the protein is kept stable in its initial conformation, for both open and closed structure of LBP, using the ELNEDIN model. In the domELNEDIN model the open conformation of LBP closes up to 3 Å and to 1.6 Å when the elastic bonds and equilibrium angles of the closed conformation are used, compared to crystal structure of the closed conformation of LBP. However, even though the domELNEDIN model introduces extra flexibility to the elastic scaffold, closed conformation does not open. Moreover, when the topology for the open conformation is applied in the simulation starting from the closed conformation, the structure keeps a stable closed conformation, although the internal domain structure and backbone changes to match those in the open conformation. This observation corroborates that the application of a topology from
2.6. PROTEIN DYNAMICS ON THE MICROSECOND TIME SCALE

a different conformation does not force a conformational change in itself. It cannot at this stage be clearly stated if the results observed from the simulations of closed conformation of LBP are biologically relevant, or model dependent.

2.6.3 Protein-Ligand Interactions

In the X-ray structure of the closed conformation, ligand is found to be bound in the cleft between the two domains by the hydrogen bonds between the backbone part of leucine and the protein (Figure 2.15A). The strongest interaction is due to the salt bridge formed with domain 2, but overall a higher number of interactions are assigned to domain 1. Additionally, ligand is stabilized in the binding cleft by non-polar interactions. The hydrophobic part of the binding cleft is formed by the side chains from a leucine, a tryptophan and a tyrosine. In the 100 ns of atomistic simulation ligand kept its position inside the protein. However, the minor rearrangements of the protein structure result in the loss of the salt bridge to domain 2 and two of the hydrogen bond interactions (Figure 2.15B). In the version 2.1 of the MARTINI force field (used in the work presented in this chapter), based on ref. 38 the side chain part of the ligand was mapped to the C1 bead type and the backbone to the P5 bead type, as this is listed as the bead type to assign the backbone of free amino acids (Table 2 in ref. 38).

Since the atomistic simulation of the closed conformation with the ligand present in the binding cleft witnessed a stable protein-ligand interaction, five domELNEDIN repeat simulations were run with this setup, where the focus was on the protein-ligand interaction and the influence of the ligand behaviour on the protein tertiary structure. The results for the five repeat simulations shows that the ligand only stays in the cleft in one simulation (Figure 2.16), but dislocates to the “bottom” of the cleft losing all binding site
CHAPTER 2. STUDYING PROTEIN DOMAIN MOVEMENTS - DOMELNEDIN MODEL

interactions. In the other four cases the ligand escapes the binding cleft and is absorbed in the bulk of water early in the simulation at around 136, 432, 548, or 846 ns of “real” time. The escape of the ligand does not seem to have influence on the overall protein structure (Figure 2.16). The escape is clearly not caused by the domain flexibility in the domELNEDIN model, as the ligand also escapes already after 236 ns of “real” time in the ELNEDIN simulation (data not shown).

Based on the results, the description of the backbone bead of the ligand was re-thought. It could be argued that the Nda bead type, which describes backbone beads in a turn, having both H-bonding donor and acceptor capabilities could be an alternative representation for the backbone part of the leucine ligand. The capability of forming hydrogen bonds from both the nitrogen and carbonyl to other molecules, are similar between a free amino acid and the backbone in a turn. When the backbone of the ligand is described by the Nda bead type, all interactions to the protein in the binding pocket will be weakened, except for the interaction with the glutamate which stays the same, compared to the ligand described with the P5 bead type. However, for Nda beads, the interaction with water is less attractive than the interactions with the protein in the binding site, whereas for P5 beads the water interactions are as attractive as the protein interaction. With the Nda-C1 description for the ligand four repeat simulations were performed, using the domELNEDIN model. However, these simulations showed similar ligand behaviour as seen for the P5-C1 mapping (data not shown) and in three out of four repeat simulations the ligand escaped from the binding cleft within the first 100 ns of simulation time while in the last case it dislocated to another position in the cleft. The only noticeable difference, compared to the simulations with the P5-C1 mapping of leucine, was that the ligand explored the surface of the protein before completely leaving for the water phase. This behaviour probably results from the mentioned difference in the preference towards water.

**Figure 2.16** - Protein-ligand distance compared to protein backbone RMSD. The distance between the Glu226 Qa bead (see Figure 2.15) and the ligand amino acid P5 bead is given for the five repeat simulations (each their own color), as well as the corresponding RMSD measured for the protein backbone.
2.7. **CONCLUSIONS**

The reason for the loss of interactions to the ligand must thus be found in the CG description of the ligand itself and its interactions with the protein residues, as protein-ligand interactions in the X-ray structure, and atomistic simulation of the closed conformation, are probably based more on directional hydrogen bonds than on the hydrophobic and hydrophilic interactions between molecules. This problem could be simply fixed by adding weak elastic bond to mimic for instance the salt bridge, as observed from the crystal structure of the closed conformation, between the backbone bead of the ligand and the Qa bead type of Glu226. In the present study we have been concerned with the description of protein structural changes, and an approach where the ligand is constrained to the binding pocket has therefore not been relevant. However, improving CG modelling of protein-ligand interactions in LBP is the subject of another project and will be described in Chapter 3.

### 2.7 Conclusions

All CG models, namely MARTINI, ELNEDIN and domELNEDIN, were applied to two structures with different levels of dynamics, the open and closed conformations of a ligand binding protein. It is shown that MARTINI clearly does not have the ability to maintain the structural integrity while the ELNEDIN extension, imposes global elastic network on the structure that keeps this structure stable. The major drawback of the ELNEDIN model is that the structure is strongly biased to the initial structure and thus has no chance of describing the protein conformational changes on the microsecond time scale. In the domELNEDIN approach, introduced in the present study, the elastic network is imposed only inside the protein domains, maintaining the structural integrity of the domains while at the same time allowing for their unrestricted movement. It is shown that even though this flexibility is introduced, the structural integrity is intact in 100 ns simulations, with protein dynamics equivalent to that observed in the ELNEDIN and atomistic simulations, but it does not result in unphysical structural changes. In microsecond long simulations, the open structure is observed to close up while the closed structure remains stable and closed. This could be assigned to experimental knowledge about the protein function and to observations from the atomistic simulations, or it could also be model dependent.

A successful modelling of protein conformational changes with the domELNEDIN model involves the ability to divide the total structure into domains for which the structure does not change much between conformations. Several methods exist that will help define these structural domains and thus an automatic workflow for the model setup can be defined, also for cases where only the structure of one conformation is known. If structures of multiple conformations are known, this can be used to establish the proper division in structural domains, and furthermore, structural information from two different conformational states give insights into the level of "induced fit" involved in the packing of the domains in the different conformations. This knowledge used in the
simulation setup can enable the modelling of a more complete conformational change, as also presented in this study. The ability to allow for “induced fit” between the domains could also possibly be achieved through a setup where instead of a uniform stiffness of the elastic network throughout the domain, the elastic network was made weaker in areas based either on distance to the center of the domain or based on knowledge from a short atomistic simulation. It has to be tested, however, if such a setup would still produce dynamics comparable to atomistic simulations on the nanosecond time scale.

As the structural changes of LBP were not affected by the release of the linkers from the elastic network bonds, it could also be worthwhile to consider loosening the regulation of the backbone angles in the linker region to further increase the linker flexibility. This could be done by lowering the force constant on the angle terms to e.g. 25 kJ·mol\(^{-1}\) (it is 40 kJ·mol\(^{-1}\) at present) as imposed on backbone beads in random coil structures in the standard MARTINI force field. Alternatively, it could be possible to achieve the flexibility of the protein backbone in the MARTINI model by re-parametrizing the dihedral angle potential, as it was done recently for the amyloid- and elastin-like peptides.\(^\text{107}\)

The present study also points to a current limitation in the MARTINI model of the interaction between the protein at study and its leucine ligand. One improvement could be to model the ligand with beads that hold particles representing more or less flexible plus and minus poles of the bead - similar to what is done in the polarizable water model constructed for MARTINI.\(^\text{44}\) This is currently incorporated in new 2.2P version of the MARTINI force field,\(^\text{108}\) and is the subject of the work presented in Chapter 3.
Chapter 3

Improving CG Modelling of Protein-Ligand Interactions

The work presented in this chapter has been performed in the laboratory of Professor Siewert-Jan Marrink at the University of Groningen, the Netherlands. The stay was financed by the HPC-EUROPA2 Transnational Access program.

Contents

3.1 Polarizable MARTINI Force Field .................................. 36
3.2 Polarizable Water Model ............................................. 38
3.3 Methods ............................................................. 39
  3.3.1 Software ......................................................... 39
  3.3.2 LBP Setups ...................................................... 39
  3.3.3 Protocol ......................................................... 39
3.4 General Corrections to Standard MARTINI Model ................. 40
3.5 Improving Protein-Ligand Interactions ............................. 41
  3.5.1 Leucine Ligand Description for Polarizable Model .......... 41
  3.5.2 Closed Conformation of LBP ................................. 42
  3.5.3 Open Conformation of LBP ................................. 46
3.6 Conclusions ......................................................... 49
The work presented in Chapter 2 showed that there were some limitations in the standard MARTINI model (version 2.1) with regard to modelling protein-ligand interactions in a leucine-binding protein. This is not the only shortcoming of the standard MARTINI model. Previous studies showed that in the standard MARTINI model the charged residues, as well as phenylalanine, proline, and to a lesser extent tryptofan, are found to be too hydrophobic. This was also confirmed by studies of the free energy of amino acid side chains dimerization in different solvents with different polarity. These studies showed that the aromatic side chains are too hydrophobic, and charged and polar interactions in a low dielectric medium are underestimated compared to the atomistic models. The binding interactions of the polar side chains of asparagine and glutamine have also been showed to be underestimated in the standard MARTINI model. The problematic residues in version 2.1 of the MARTINI model have been re-parametrized, and a new set of parameters has just recently been released and denoted by 2.2 or 2.2P to be used in combination with standard or polarizable water model, respectively.

3.1 Polarizable MARTINI Force Field

In version 2.2 new bead types were assigned for proline (Pro), phenylalanine (Phe), and tryptophan (Trp) side chains to improve partitioning free energies. For the Phe side chain the improvements were achieved by using slightly more polar bead types for the ring beads switching from SC4 to SC5. For the Pro side chain the original assignment of beads was changed from Na-C2 to P4-C3, arguing that the backbone polarity should be less than that of a normal side chain due to the reduced H-bonding propensity, and that the side chain analogue is actually like a propane and therefore should be kept rather apolar. This applies only to the case where the residue is part of an unstructured chain, otherwise when being a part of an α-helix or a β-strand, less polar bead types are used. The original parametrization of Trp ring system (SC4-SC4-SC4-SP1) has been replaced by SC5-SC5-SC4-SPd, which better reflects the chemical nature of the Trp side chain, which now has hydrogen bond donor capabilities (Sn). In general the bonded terms were adjusted to improve the length of the standard α-helices and increase numerical stability for poly-alanine and glycine repeats.

The main improvement of the new version was the introduction of an "off-centre charge" model, which allowed for more realistic mimicking of the interactions between pairs of oppositely charged residues, which have been underestimated in 2.1 version, and for the improved dimerization in apolar environment and inter-facial binding for polar side chains (see Figure 3.1). In the standard MARTINI model, the outermost side chain bead of the charged residues would be denoted as a Q bead type, and therefore it would be able to interact with other charged beads via Coulomb interactions. In the new 2.2P version this bead (still denoted as Q bead type) only interacts with other beads via van der Waals interactions, as a new particle is bearing the charge at a distance of 0.11 nm from the centre of the Q bead type, and this particle interacts with other charged beads.
3.1. POLARIZABLE MARTINI FORCE FIELD

Figure 3.1 – Schematic drawing of a Ser-Glu dipeptide showing the new “off-centre charge” model for charged and polar amino acids. The charged Glu residue is represented by a backbone bead (BB) and one side chain bead (SC1) which interacts with other beads via van der Waals interactions. An additional particle S1n is constrained at 0.11 nm from the centre of the SC1 bead, and carries the full negative charge of Glu but has no van der Waals interaction. The polar Ser residue is represented by a backbone bead (BB) and a side chain bead SC1 that carries the van der Waals interaction. It is a virtual site with the centre in the middle of two additional particles S1n and S1p that carry a negative and a positive partial charge, respectively. They are constrained at a distance of 0.28 nm and they do not interact via van der Waals interactions.

via Coulomb interactions. For instance the glutamate (Glu) (Figure 3.1) side chain bead SC1 (which is of type Q) has an extra particle S1n (that is a D bead type, which stands for dummy bead). This particle carries the full negative charge of the Glu residue. It is bound to the centre of SC1 bead using a 0.11 nm constraint. The van der Waals centre (side chain bead - SC1) and the charged bead (particle S1n) both have equal mass of 36 amu, which is half of the mass of a standard bead (72 amu). With the new off-centred position of the charge, two charges from different beads may approach each other much more closely, increasing the interaction by approximately a factor of 3 for oppositely charged amino acids compared with version 2.1 of the MARTINI model,\textsuperscript{38} which is in better agreement with the atomistic model.\textsuperscript{98} In case of polar residues like serine (Ser) (Figure 3.1), the side chain bead SC1 is now a virtual site (does not have mass), and it is involved in the van der Waals interactions with other particles in the system. The SC1 bead is positioned in the middle of two additional particles, S1n and S1p, which are constrained at a distance of 0.28 nm from each other and bear a negative and a positive partial charge, respectively. Due to the constraint the polarized residues have a fixed dipole moment, that can contribute to orientational polarization. The angle between the partial charges and virtual site remains fixed at 180 degrees. The strength of the Lennard-Jones potential, due to the addition of explicit particles, is reduced compared to the standard MARTINI model. The partial charges are real sites, each of mass of 36 amu, that do not have van der Waals interaction, but are interacting via Coulomb potential with other charged particles. The polarizable MARTINI model has to be used with the polarizable water model\textsuperscript{44} that is introduced in the next section.
3.2 Polarizable Water Model

In the standard MARTINI water model\textsuperscript{37,38,42} as described in Chapter 1, the water bead models four water molecules, and is described by the P4 bead type. The standard CG water beads do not have any screening capabilities, as they do not bear charges. Therefore, the dielectric constant is set up to $\epsilon_{rel} = 15$. This together with the smooth shifting of the electrostatic interaction towards the cut-off (1.2 nm) allows for distance dependent dielectric screening. However, this is a quite unphysical situation, and thus a new water model for MARTINI was introduced, where additional particles bearing charges were placed on the water beads.\textsuperscript{44} The water bead type was changed to a new POL bead type with a mass of 24 amu, and it interacts with other beads via van der Waals interaction. The additional charged particles denoted as WP and WM bear negative and positive charge of $q = \pm 0.46$, respectively (Figure 3.2), also both of them have mass of 24 amu. They are constrained to the central bead at the distance of 0.14 nm, as thus as in the polarizable protein model, they are inside the van der Waals radius, preventing clashes between charged beads. As opposed to the polarizable residues, WP and WM are not constrained with each other. Therefore, the charged particles can rotate with respect to each other, allowing the dipole moment to change. Thus, in an apolar environment the water bead has an effective charge of zero, as the charged WP and WM particles collapse onto each other, because the equilibrium angle $\theta = 0^\circ$, while in a polar environment the water beads achieve an effective dipolar moment. The WP and WM particles of the same water bead do not interact with each other, as there is exclusion put on them. Furthermore, they interact with other charged particles via Coulomb potential. The introduction of the charged particles in the water bead allowed the mimicking of the strong polarization and charge behaviour of real water molecules in a more realistic manner, and the relative dielectric constant of the system could be lowered to $\epsilon_{rel} = 2.5$.

![Schematic drawing of a polarizable water bead.](image)

Figure 3.2 - Schematic drawing of a polarizable water bead. The bead is represented by central POL bead carrying the van der Waals interactions, and two additional particles WP and WM, that carry a positive and a negative partial charge, respectively. They interact with other charged beads/particles via Coulomb interactions. The equilibrium angle between WP and WM particles is denoted as $\theta$.

The addition of charged particles both, for the side chains of protein and on the water beads, makes the simulations more computationally expensive than the simulations using version 2.1 of the MARTINI with the standard water model\textsuperscript{37,38,42} The computation is slowed down by up to a factor of three.\textsuperscript{44}
3.3 METHODS

3.3 Methods

This project investigates the possibility of improving the description of protein-ligand interactions in a periplasmic leucine-binding protein (LBP), using the new MARTINI model where additional charged particles on the polar residues were introduced. The same LBP structures were used as the test examples for the new version of MARTINI model, as in Chapter 2.

3.3.1 Software

The GROMACS package version 4.5.5\textsuperscript{72,73} was used for all simulations, using the MARTINI force field version 2.2 and 2.2P,\textsuperscript{98} with the standard\textsuperscript{37,42} and the polarizable water model,\textsuperscript{44} respectively. The domELNEDIN model was used to stabilize the tertiary structure of the protein in different setups. Plots were prepared using Xmgrace.\textsuperscript{110}

3.3.2 LBP Setups

Four different setups were built starting from a closed (PDB: 1USK\textsuperscript{51}) or an open (PDB: 1USG\textsuperscript{51}) conformation of LBP. For the closed structure two different setups were applied; either with or without the ligand present in the binding cleft. For the open structure also two different setups were applied; either without ligand or with ligand positioned in the close proximity of Glu226 of the protein, in the binding cleft, to form a salt bridge, as observed in the crystal structure of closed conformation of LBP.

3.3.3 Protocol

Each system was coarse-grained using similar protocol as described for previous systems in Chapter 2, however using either the 2.2 or 2.2P version of the MARTINI model. The prepared protein structures were placed in a cubic box with dimensions of 100 Å. Proteins were energy minimized in vacuum to relax the structure. The counter ions were added (9 Na\textsuperscript{+}), followed by the addition of water beads; standard water beads\textsuperscript{37,38,42} for 2.2 version and polarizable water beads\textsuperscript{44} for 2.2P version of the MARTINI model. The systems were again energy minimized and position restraints were applied for 1 ns using 20 fs time step, to relax water and ion beads around protein. Production runs were of 200 ns for 2.2 version and 1 \(\mu\)s for 2.2P version, for both using the time step of 20 fs. For better sampling some of the simulations were extended to 2 \(\mu\)s, if so it is stated in the text. Some of the setups included ten ligands, if so, the ligands were positioned randomly in the simulations box, before solvation, and protocol remained the same as described. On the contrary to the previous Chapter 2, the time given in this chapter is a CG time and not “real” time.
CHAPTER 3. IMPROVING CG MODELLING OF PROTEIN-LIGAND INTERACTIONS

3.4 General Corrections to Standard MARTINI Model

First to see whether the changes of Pro, Phe, and Trp side chains to different bead types, and the adjustment of the bonded terms are sufficient for improving the protein-ligand interactions, five repeat simulations were set up, starting from the closed conformation of LBP with the ligand bound, using the version 2.2 of MARTINI model with the standard water model. Simulations were run for 200 ns. The results (Figure 3.3A) show that in three out of five cases leucine breaks the interaction to Glu226 at 4.5 ns (blue line), 30 ns (orange line), or 78 ns (black line). It has to be pointed out that in the third simulation, where ligand escaped at 78 ns of simulation time, starting from ≈40 ns protein-ligand interactions are lost and retrieved until the final escape from the binding cleft. In two other simulations, where ligand escaped, the blue and orange lines at the distance of ≈1.1 nm (Figure 3.3A) indicate that the ligand is still positioned in the binding cleft, however close to the water phase. In the remaining two simulations protein-ligand interactions are stable during the simulation (green and red lines). The corresponding histogram (Figure 3.3B) shows that when the centre of mass of the backbone bead of ligand (P5 bead type) is at the distance of 0.6 nm from the centre of mass of the side chain bead of Glu226 (Qa bead type) it forms a stable salt bridge. When the distance increases to ≈1.1 nm the protein-ligand interactions are lost, however, the ligand is still in the proximity of the protein. Above ≈1.3 nm the ligand moves into the water phase.

Figure 3.3 – Left: Protein-ligand distance compared to protein backbone RMSD. The distance between the protein and ligand corresponds to the salt bridge between Glu226 and leucine. In the CG simulations the distance is measured between the side chain bead of Glu226 (Qa bead type) and the backbone bead of the ligand (P5 bead type). The distance is given for the five repeat simulations (each their own color), as well as the corresponding RMSD measured for the protein backbone. Right: Histogram showing the distribution probability of the protein-ligand distances shown on the Figure on the left, with corresponding colors. The histogram shows that at the distance of 0.6 nm protein-ligand interactions are lost, however at the distance ≈ 1.1 nm the ligand is still in the binding pocket closed to the escape.
3.5 Improving Protein-Ligand Interactions

The introduced corrections to the bonded terms do not directly influence the binding mode of the ligand, however, there is an improvement compared to 2.1 version of the MARTINI model, in which the interaction mimicking the salt bridge was lost in all studied cases. However, as these results are not repeatable among five performed simulations no further discussion on the 2.2 version of the MARTINI model will be given. Instead the focus will be on the 2.2P polarizable version.

3.5 Improving Protein-Ligand Interactions

The new “off-centre charge” model introduces the additional charges on the polar residues, and shifts the charge 0.11 nm outwards in the side chain beads for the charged particles. This should allow for more realistic modeling of the contact pairs of oppositely charged residues, and potentially has an influence on the protein-ligand binding, because now the Glu226 could approach the ligand charge at a closer distance, which can result in the formation of a stable salt bridge. The charge on the ligand would interact via Coulomb potential, which is a longer-range interaction compared to the van der Waals interactions. Additionally, contributions from polar residues, like Ser79, that now bear additional partial charged particles on the side chain beads should give a significant contribution to stabilizing ligand in the binding mode, compared to 2.1 version of the MARTINI model. However, for this to be true, the ligand description also has to be changed.

3.5.1 Leucine Ligand Description for Polarizable Model

Generating protein topology for polarizable version of the MARTINI model is done using martiniize.py script (it can be found on the MARTINI home page). This program converts an atomistic structure to a CG structure, yielding corresponding CG topologies, however, it does not describe how to properly map small free molecules like the leucine ligand in our system. Thus, a description for the leucine ligand compatible with the polarizable force field for proteins was made (Figure 3.4).

![Figure 3.4](image_url) - Schematic description of a free leucine ligand with a polarizable backbone bead. The backbone bead (BB) is represented by a central bead that carries the van der Waals interaction, and two additional charged particles: Bp and Bn that carry a positive and a negative charges, respectively. The charged particles interact via Coulomb potential. A side chain bead (SC1) remains unchanged.
In the standard CG representation the free leucine was described by P5-C1 mapping, for backbone bead (BB) and side chain bead (SC1), respectively. This was shown to be an insufficient description of ligand to form a stable interactions with the protein. In the new representation the BB is still of P5 type, however, it is now modeled by a polarizable bead bearing charges (Figure 3.4). The two additional charged particles Bp and Bn, bear a full positive and negative charge, respectively, mimicking the amino and carboxyl groups of a free leucine. These particles have a mass of 36 amu, and are constrained between each other at the distance of 0.22 nm, as it was done for the charged particles in the version 2.2P of the MARTINI model. To prevent back flip of charges additional angle was defined between Bn-BB-SC1 beads with the value of 90° and a force constant of 100 kJ·mol⁻¹. The introduced charges interact with other charged particles only via Coulomb interactions. The C1 bead type used for the side chain bead remained a standard C1 bead type.

3.5.2 Closed Conformation of LBP

As described in Chapter 2 the X-ray structure of the closed conformation contained ligand bound in the cleft between the two domains. It was held in the place, by the hydrogen bonds between the backbone part of the leucine ligand and the protein. The strongest interaction is due to the salt bridge formed with domain 2, but overall most interactions are assign to domain 1 (Figure 3.5A). As mentioned before, in the CG description of the standard MARTINI model, the protein-ligand interactions are modeled only based on the polarity of the interacting beads (Figure 3.5B). To see if the changes presented in 2.2P version of the MARTINI model i.e. shifting a charge for the Glu226 from the centre to the edge of a charged bead, introducing plus and minus partial charges for polar residues like Ser79, and Thr102 (Figure 3.5C), and changing the description of the ligand, improve the modelling of protein-ligand interactions five repeat simulations were run.
The simulations were set up starting from the closed conformation of LBP with ligand bound in the binding cleft, and run for 1 µs using 2.2P version of the MARTINI model, along with the polarizable water model.\textsuperscript{44} The results are shown in Figure 3.6.

**Figure 3.6** - Protein-ligand distance compared to the protein backbone RMSD for the repeat simulation #1 - #5. The distance was measured between the COM of the side chain bead S1 (Qa bead type) and charge particle S1n (D bead type) of the Glu226 (red line) or the side chain bead S1 (N0 bead type) of Ser79 (blue line) to the COM of the backbone bead BB (P5 bead type) of the ligand leucine. The corresponding RMSD measured for the protein backbone is also included in the figure (black line).
In all cases the salt bridge formed between the Qa bead type of Glu226 and the bead type P5 of the ligand is stable and maintained during the simulation. From the same plots it is seen that during the simulations the ligand interacts almost constantly with the polar residue Ser79, however, the interaction is not as strong as in the case of Glu226. This is due to the significant difference in charges. The Ser79 have a partial charge equal to $q = \pm 0.4$, and the Glu226 bears a full positive charge $q = 1$. When investigating further the remaining interactions between the ligand and the protein (see Figure 3.5C) it is observed that also Thr102 forms stable interactions with the ligand at a distance 0.1 nm higher than for Glu226. During all five simulations, the protein-ligand interactions were kept and the overall protein structure was stable and remained in the closed conformation during the simulation, however, with the exception of a 50 ns period for repeat simulations #3 and #4 in which the structure opens slightly. In simulation #4 the structure returns back to its initial conformation after 50 ns, while for simulation #3 it is not possible to state what happens further, as the opening occurred at the end of the simulation (Figure 3.6).

The introduction of the “off-centre charge” model, and the introduction of additional charged particles for side chain beads of polar residues, as well as for the backbone bead of the leucine ligand, show that with the new 2.2P version of the MARTINI model it is possible to study protein-ligand interactions for LBP, and producing stable conformations. To see how specific the interaction between the leucine ligand and LBP is, several other tests were performed, and will be presented in the following sections.

**Removing Ligand from the Binding Cleft**

To see whether it was possible for the ligand to enter the binding cleft of the closed conformation of LBP, a simulation was run, starting from the closed conformation of LBP, without the ligand in the binding cleft. Ten leucine ligands, however, were randomly positioned in the simulation box and the simulation was run for 1 $\mu$s. Due to the tight packing of the domains relative to each other, the ligands were not able to enter the binding cleft. This was expected since the entrance to the binding cleft is too small for the molecule to pass through. Additionally, the behaviour of ligands was investigated, and it was found that ligands do not aggregate in the water phase, however they are able to bind to the charged particles at the protein surface and form aggregates up to four molecules at a time. These aggregates are able to disassociate during the simulations. To see if the removal of the ligand influenced the protein structure RMSD for backbone beads was measured, showing that the overall structure was stable at the level of 3 Å. The tertiary structure of the protein was kept stable during the simulation using domELNEDIN and it was therefore possible that a lack of ligand in the binding pocket would induce the opening of the protein. However, as in the case of the standard MARTINI model (Chapter 2), no opening towards the known open conformation was observed, suggesting that this could be biologically relevant or model dependent behaviour.
3.5. IMPROVING PROTEIN-LIGAND INTERACTIONS

Pulling Domains Apart

To study how the ligand affects and mediates the interaction between the domains, simulations where the domains were pulled apart were set up. Five repeat simulations of the setup starting from the closed conformation of LBP with the ligand present in the binding cleft, as well as five repeat simulations of the setup starting from the closed conformation without the ligand present in the binding cleft, were run for 200 ns.

The pulling code used in the GROMACS package applies forces or constraints between the centers of mass (COM) of a reference group and one or more pulling groups. In the simple case, as the one presented in this subsection, there is one reference and one pull group that are treated the same. The reference group is domain 1 and the pulling group is domain 2. Three different types of calculation are supported, i.e., umbrella pulling, constraint pulling, and constant force pulling. In this case, the constraint pulling was chosen, and thus domain 1 and 2 were pulled apart using rigid constraint. The rate of change of the reference position was set to 0.01 nm per ns.

The results are shown in Figure 3.7, and it is seen that there is very little difference between pulling apart domains of the protein with and without ligand. With a very high standard deviations for the force needed to pull the protein domains apart (774.2 kJ·mol\(^{-1}\)·nm\(^{-1}\) for system with the ligand, and 765.1 kJ·mol\(^{-1}\)·nm\(^{-1}\) for system without the ligand), the pull force is similar for both LBP systems, meaning that the ligand affects and mediates the interactions between domains to a very little extent.

![Figure 3.7](image-url) - The pull force as a function of the distance, given for the setups of the closed conformation with and without the ligand. The values are averaged over the results from the five repeat simulations for each system. The black line indicates the setup with a ligand bound, and the red line indicates the setup without a ligand bound in the binding cleft.
3.5.3 Open Conformation of LBP

To see whether it is possible for the ligand to find its proper binding mode, when positioned far from the protein, simulations starting from open conformation of LBP with ten randomly distributed leucine ligands in the simulation box were set up.

Random Positioning of Ligands

The first simulation was run starting from the open conformation of LBP with ten ligands randomly positioned in the simulating box. The result of 1 µs simulation (Figure 3.8) was that only one time a ligand went into the binding cleft with the result of a salt bridge formation between the side chain bead of Glu226 and the backbone bead of the ligand. In this simulation the charged particles approached each other closest at distance of 0.25 nm.

The ligand binding does not seem to have a significant influence on the overall structure of the open protein, since the closing of the protein is initiated already at the beginning of the simulation. The ligand binds at ≈25 ns followed by disassociation of the ligand, which moves towards the exit of the binding pocket followed by binding again and finally the ligand escapes at ≈170 ns. The protein closes with the final RMSD value of 4 Å compared to the closed crystal structure. For the clarity the conformations that display incomplete closing will be referred to as semi-open conformations in the following sections.

Forming a Salt Bridge

The spontaneous binding of a ligand to the protein, leads to the conclusion that if the protein is stabilized by a ligand it would close more towards the already known closed conformation. It was interesting to see whether the ligand did form a salt bridge to the protein at the beginning of the simulation and can it stay there during the simulations.
or if it would be replaced by another ligand that can enter the cleft of the semi-open conformation. For these reasons three repeat simulations were set up starting from open conformation of LBP with one ligand positioned in the binding cleft to form a salt bridge with the protein, and nine additional ligands distributed randomly in the simulating box. Simulations were run for 1 $\mu$s and were further extended to 2 $\mu$s for better sampling. The results are shown in Figure 3.9.

![Figure 3.9](image)

**Figure 3.9** - Protein-ligand distances between the COM of the side chain bead particles ($Q_a$ and $D$ bead types) for Glu226 to the backbone bead of leucine ligand ($P_3$ bead type), shown for the repeat simulations #1 - #3. For the ten ligands the measured distance is shown which their own color.

When ligand interacts with the Glu226 with its full positive charge the negative charge points directly towards the exit of the binding cleft. This makes it easier for other leucine ligands to bind to the negative charge with their positively charged particle. Moreover, as leucine is hydrophobic it preferably moves towards the hydrophobic cleft of the protein. This leads to the situation where one of the ligands is bound to the Glu226 forming a salt bridge, while more ligands are found in the binding cleft interacting with each other. This situation is visible especially for repeat simulation #2 (Figure 3.9), where at distance of 0.25 nm three different leucines are found to exchange positions with each other. Additionally, at the distance starting from 0.5 nm to 1.0 nm, from the Glu226, there are up to four leucine ligands interacting with each other at a time via Coulomb
potential. In that way they form linear aggregate stabilized by the interactions between their full charges and the hydrophobicity of the binding pocket.

When looking closer at the snapshots from the repeat simulation \#2 (Figure 3.10) it is seen that one ligand interacts with its positive charge with the negative charge of Glu226 (forming a salt bridge), and the ligand also interacts with its positive charge with the negative charge of another ligand positioned in the binding cleft. At the same time there are two additional ligands present in the binding cleft close to the ligand which is interacting with protein. When one of these ligands moves closer to Glu226, the ligand that initially formed a salt bridge loses its interaction to Glu226, and the new ligand gains the interaction with Glu226. The replaced ligand moves out between the remaining ligands forming a linear aggregate with them. This behaviour is to some extent similar to the “knock-on” mechanism proposed for the potassium channel by Hodgkin and Keynes,\textsuperscript{112} in which translocation of two selectivity filter-bound ions is driven by a third, incoming ion.

![Figure 3.10](image-url)

**Figure 3.10** – “Knock-on” mechanism. The shown snapshots are given for frames around 0.5 µs of simulation \#2, with the colors of the ligands corresponding to the colors given in Figure 3.9. The backbone of the protein is shown in Licorice representation in gray and Glu226 residue is shown in yellow. (A) The knock-on mechanism starts when the green ligand interacts with Glu226 (forming a salt bridge) and the ligand interacts with its positive charge with the negative charge of the brown ligand. At the same time the red and the blue ligands are positioned in the binding cleft. (B) The red ligand moves closer to Glu226. (C) In next step the red ligand replaces the green ligand, which now mediates the interactions between the blue and brown ligands. (D) The blue, green, and brown ligands form a linear aggregate, while the red ligand interacts with Glu226.

The overall structure of the protein is changed towards the closed conformation with the final RMSD value of 3 Å, 4 Å, 4 Å, for repeat simulations \#1, \#2, \#3, respectively. To test whether the mode with more than one ligand in the binding cleft hinders the protein from closing up completely, since the ligands occupy the space that potentially is
needed for the domains to move towards each other, new simulation was set up. For the starting structure, the structure as it occurs at the 1 $\mu$s of the repeat simulation #1, was used. All leucine ligands, besides the one forming a salt bridge, were removed and the simulation was run for another 1 $\mu$s. No further closing of the protein was observed, and the RMSD value for the backbone of the protein remained at the level of 3 Å compared to the closed crystal structure of LBP.

In all previous simulations counter ions were used to neutralize the system, and no aggregates in the water phase were observed. To see if the ionic concentration could provoke leucine ligands to aggregate in the solvent, 0.2 M of NaCl was added to the system. The results, were comparable to the previous simulation of the closed conformation with ten randomly positioned ligands in the box, where no ligand aggregates were found in the water phase. As seen before aggregates where only found in the protein cleft or when bound to the protein surface.

3.6 Conclusions

The new polarizable MARTINI force field has shown significant improvement compared to the previous version 2.1 of the MARTINI model, in the ability to reproduce ligand binding to the LBP. It has been shown that when the protein is in a closed conformation it forms a stable salt bridge between Glu226 and the ligand which also interacts strongly with residues from domain 1. When the ligand is removed from the closed conformation of LBP and placed randomly in the box the binding cavity protected by the two domains is too narrow for leucine ligand to enter. On the contrary when the open conformation is presented without the ligand bound the randomly positioned ligands can find the correct binding mode. At the beginning of the simulation, the ligand forms a salt bridge to the open structure of the protein, and hence more binding events are observed, including exchange of ligand in the binding cleft, and formation of linear aggregates in the cleft and at the surface of the protein. These aggregates have the ability to disassociate and reassociate during the simulation. Moreover, the aggregation of the ligands was not observed in the water phase even with the significant increase of ion concentration. This MD study shows that introducing charges on the side chain beads of polar residues, off-centre charge on the side chain beads of charged residues and charged representation of ligand improved the protein-ligand interactions in the case of LBP.

The results concerning the transition between the open and the closed conformations of LBP where within the 2.2P version of the MARTINI model comparable to the results observed in the simulations using the standard MARTINI model (described in Chapter 2). The simulations of the closed conformation of LBP showed a stable closed protein structure, while the simulations of the open conformation of LBP showed that this structure can close up at the lowest to RMSD value of 3 Å, compared to the closed crystal structure. The reason for the closing of the open structure, did not seem to be induced by a ligand binding, nevertheless, it can lay in the fact that in the crystal structure used
for the simulation starting from an open conformation LBP is packing as a dimer, which could stabilize this very open conformation, and a structural change to a semi-open conformation in solution could be expected.

The new model introduces polarizable particles only for the side chain beads, and thus the secondary structure still has to be imposed, which seems to have an influence on the dynamics of LBP. One way of overcoming this limitation would be to use a similar approach as for polarizable particles, namely adding partial charged particles to the backbone beads of the protein to mimic the hydrogen bonding of secondary structure elements. Another solution that could be transferred to proteins, is a method that has been recently presented for the amyloid- and elastin-like peptides.\textsuperscript{107} In this model the backbone of the peptide is flexible during the simulation, thanks to the re-parametrization of the dihedral angle potential, which was fitted to the potentials of mean force derived from atomistic simulations.
Chapter 4

SERCA and Cholesterol

This chapter describes the project concerning the regulatory function of cholesterol on the SERCA. The project includes self-assembly of a lipid-cholesterol bilayer around SERCA using the MARTINI model to investigate possible binding spots for cholesterol on the transmembrane surface of SERCA.

Contents

4.1 Ca$^{2+}$-ATPase (SERCA) ............................................ 52
  4.1.1 The SERCA Membrane Environment ....................... 54
  4.1.2 Annular vs. Non-annular Binding Sites .................... 54
  4.1.3 Cholesterol as a Regulator ................................. 54

4.2 Methods .............................................................. 56
  4.2.1 Protocol ......................................................... 56
  4.2.2 Simulation Parameters ....................................... 57
  4.2.3 Analysis ....................................................... 57

4.3 Results and Discussion ........................................... 58
  4.3.1 Bilayer Self-Assembly ....................................... 58
  4.3.2 Transmembrane Surface of SERCA ......................... 59
  4.3.3 Possible Interaction Spots for Cholesterol ............... 59
  4.3.4 Annular Interactions ....................................... 61
  4.3.5 Potential Non-annular Binding Interactions .............. 63

4.4 Perspectives ....................................................... 67

4.5 Conclusions ....................................................... 68
CHAPTER 4. SERCA AND CHOLESTEROL

4.1 Ca$^{2+}$-ATPase (SERCA)

The control and maintenance of electrochemical gradients in cells and across their membranes is crucial for living cells. The P-type ATPase superfamily comprises transmembrane ion pumps which are key players in maintaining ion gradients through primary, active transport. The SERCA (Figure 4.1) is the structurally most well-studied ATPase. It consists of 994 amino acid residues organized in a transmembrane (TM) domain consisting of ten TM $\alpha$-helices and three cytoplasmic domains. The cytoplasmic domain consists of three sub domains, the actuator (A) domain is involved in the coupling events between the TM $\alpha$-helices and the cytoplasmic domains. The phosphorylation (P) domain, holds an aspartic acid residue (Asp351) that undergoes covalent phosphorylation during the functional cycle. The nucleotide binding domain (N) is responsible for binding ATP, allowing the $\gamma$-phosphate of ATP to be transferred to Asp351 during the hydrolysis of ATP.

Figure 4.1 – Structure of SERCA in the “E2” state (PDB: 2C8K). The A-domain is shown in yellow, the N-domain in red, the P-domain in blue, TM1-2 in magenta, TM3-4 in green, TM5-6 in orange, TM7-10 in iceblue, and loop regions in gray. The position of the SERCA inhibitor thapsigargin is marked with an orange triangle in the TM part of SERCA. The cytoplasmic and SR lumen sites are indicated in the picture.
SERCA was one of the first eukaryotic membrane transporters for which the 3D structure and functional cycle were clarified using X-ray crystallography. The first structure of SERCA was determined in 2000. To this day there are more than 30 structures of SERCA in different functional states including complexes with different inhibitors.

The ion transport mechanism of SERCA is seen in Figure 4.2. During the cycle SERCA moves between two major states; “E1” which presents a high affinity for calcium ions and low affinity for protons at the cytoplasmic side, and “E2” which has a low affinity for calcium ions and high affinity for protons at the sarcoplasmic reticulum (SR) lumen side. The functional cycle starts when two Ca\(^{2+}\) ions bind to SERCA in the non-phosphorylated “E2” conformation with bound ATP. This happens in exchange of 2-3 protons, leading to a conformational change of the ATPase to the 2Ca\(^{2+}\)E1-ATP intermediate state (Figure 4.2 top left corner). Thereafter, a rapid phosphorylation of Asp351 occurs, resulting in the formation of the [2Ca\(^{2+}\)]E1-P:ADP intermediate state with occluded Ca\(^{2+}\) ions and bound ADP. The luminal part of the SERCA TM domains opens, accompanied by deocclusion. The bound Ca\(^{2+}\) ions are translocated across the membrane leading to a 2Ca\(^{2+}\)E2P-ATP state in which the Ca\(^{2+}\) ions are exchanged with 2-3 protons. Next by the closing the luminal channel the [2-3H \(^{+}\)]E2P-ATP is formed, with protons occluded inside the membrane helices. Finally, the latter intermediate is dephosphorylated, leading to a formation of 2-3H \(^{+}\)E2-ATP state presenting protons to the cytosolic side, and thus the transport cycle is completed.

SERCA can be inhibited, and prevented from adopting a conformation in which it can bind Ca\(^{2+}\) ions. Inhibition of SERCA leads to elevated cytoplasmic Ca\(^{2+}\) ion concentration which is cytotoxic and can trigger apoptosis. The inhibitor thapsigargin binds to SERCA in a cleft between TM α-helices M3, M5 and M7, on the lipid-exposed surface of SERCA (Figure 4.1). By keeping the TM α-helices M3 and M7 apart, thapsigargin prevents the ATPase from adopting a conformation in which it can bind Ca\(^{2+}\) ions and thus be active.
4.1.1 The SERCA Membrane Environment

Biological membranes are composed of different lipids, of which the major lipids present in eukaryotic membranes are phosphatidylcholine (PC), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidic acid (PA), phosphatidylinositol (PI), and phosphatidylserine (PS). The specific composition and physical properties of a membrane do influence the biological function and activity of some membrane proteins. In the case of SERCA, the membrane thickness and the lipid head groups are known to play an important role. The highest activity of SERCA is supported by dioleoyl PC (DOPC, di(C18:1)PC) bilayers, and the activity decreases for both PC with longer or shorter fatty acyl chains. Embedding SERCA in bilayers of dimyristoyleoyl PC (DMPC, di(C14:1)PC) or dinervonoyl phosphatidylcholine (di(C24:1)PC) lowers the activity of SERCA, and thus instead of two Ca$^{2+}$ ions, one Ca$^{2+}$ ion is bound per ATPase molecule. On reconstitution to dilauroyl PC (DLPC, di(C12:0)PC) even lower transport activity is observed most likely caused by the fact that this membrane is very leaky to Ca$^{2+}$ ions.

In vivo, SERCA is located in the endoplasmic reticulum (ER)/SR membrane, which is composed of approximately 70% PC, 23% PE and 7% PS and PI. 80-90% of the acyl chains, both saturated and unsaturated, originate from fatty acids with tails of either 16 or 18 carbon atoms.

4.1.2 Annular vs. Non-annular Binding Sites

From electron spin resonance (ESR) experiments with spin-labeled lipid bilayers, it is apparent that, in the presence of protein, a group of lipids exists which are motionally restricted compared to the bulk lipids. This group of lipids adapts to the rough surface of the protein and surrounds it by forming a shell or annulus equivalent to the solvent layer surrounding a water-soluble protein. The lipids in the annulus are not distinguishable from the bulk lipids by NMR techniques, and the exchange between the annulus and the bulk is thus fast on the NMR time scale. The interaction between annular lipids and the protein surface is in general unspecific, and merely serves as the proper solution environment to the hydrophobic surface of the protein. However, several cases exist where e.g. anionic lipids or cholesterol stabilizes the structure of the protein in the membrane or regulating its activity. Membrane components interacting with the protein in such a specific manner are termed non-annular binders and are typically found in between protein subunits or in cavities formed by the TM α-helices.

4.1.3 Cholesterol as a Regulator

Even if lipids build up the matrix of cellular membranes, sterols are essential components of these membranes. In contrast to the diversity of lipids, mammalian cells contain one major sterol, cholesterol, which is absolutely required for viability, however, it can be lethal for the organism as a whole. Most of the cholesterol is found in the
plasma membrane, and depending on the cell type, and the assay, plasma membranes have been reported to contain between 40% and 90% of the total cellular unesterified cholesterol.\textsuperscript{132-134} Even though cholesterol biosynthesis takes place mainly in the ER, the ER as well as mitochondria have a very low cholesterol content, whereas the Golgi compartment contains intermediate amounts of cholesterol.\textsuperscript{131,135}

An important function of cholesterol is its ability to modulate the physicochemical properties of cellular membranes. By interacting with membrane phospholipids, cholesterol influences their behaviour and can increase the bilayer thickness of membranes consisting of phospholipids with tails up to 16 carbon atoms.\textsuperscript{136} Because cholesterol affects membrane properties, it can also manipulate the behaviour and functions of proteins residing in the membrane,\textsuperscript{137} including ion channels, membrane receptors and enzymes, that are sensitive to physical changes in the surrounding lipid bilayer e.g. cholesterol modulate Na\textsuperscript{+}/K\textsuperscript{+}-ATPase activity.\textsuperscript{138}

Since the early 1970s, it has been pondered over whether cholesterol not only has an indirect effect on SERCA through the membrane environment or thickness, but also a direct effect. Lee and co-workers studied competition between sterols and phospholipids for binding sites at the phospholipid-protein interface of SERCA using fluorescence quenching properties of brominated lipid-cholesterol derivative and cholesterol analogs,\textsuperscript{121,139} in order to address this question and dissect whether an observed effect of a particular lipid component arises from the bulk membrane properties or from specific interactions of sterols on SERCA. Series of experiments show that lipid and cholesterol interactions with SERCA were almost identical for all the studied sterols.\textsuperscript{121,139} In general, the titration of sterol into a bilayer consisting of brominated lipids did not alleviate the tryptophan quenching exerted by the lipids, which showed the inability of the cholesterol molecules to exchange with the lipids at the surface of the protein. However, if brominated sterol was added instead, the quenching was more efficient, which lead to the suggestion that while the sterols do not exchange with the lipids in the annulus to a considerable degree, they can bind in non-annular sites, where the brominated lipid cannot go. When the regular lipid bilayer was titrated with brominated sterol, a quenching of the same efficiency as seen for the brominated lipids was achieved, for a molar ratio between the sterol and the lipid of approximately 1:1. Their conclusion was that there must be non-annular sites on the TM part of SERCA where cholesterol can interact directly.\textsuperscript{121,127,139-141} This direct interaction was proposed to happen before the structure of SERCA was even known.\textsuperscript{139}

Today, when more than 30 structures of SERCA are known, it is speculated whether a non-annular site of cholesterol could indeed be overlapping with the thapsigargin binding pocket.\textsuperscript{142,143} A number of crystal structures in which PE and PC lipids trapped in cavities on the TM surface have been modeled,\textsuperscript{144} and the presence of non-annular sites specific for cholesterol has been suggested to explain effects not compatible with competitive binding of cholesterol and lipid to the annular sites around the ATPase.\textsuperscript{139}
In this study we aim to understand the cholesterol-SERCA interaction from a structural perspective. CG simulations have been previously used to predict lipid-protein interactions for a wide range of membrane proteins.\textsuperscript{36,145,146} CG simulations allow the dynamics self-assembly of lipids around a protein of interest, enabling prediction of its position and orientation within a membrane.\textsuperscript{36,145} However, while CG simulations allow sampling of protein-lipid interactions and a “dynamic docking” of the membrane molecules to the protein surface, the description of the detailed atomistic interactions is lacking. To alleviate this, the CG information on protein-lipid interaction can be translated into atomistic resolution, for further study of the interactions on the nanosecond time scale.\textsuperscript{147–149} This approach can be referred to as a serial multiscale simulation, where the serial change of model/setup resolution takes advantage of both the time scale accessible by the CG method and the details accessible by the atomistic method. Thus, this allows both for a prediction and subsequent refinement, and analysis of protein-cholesterol interactions.

### 4.2 Methods

To search for cholesterol binding spots on the TM surface of SERCA, 30 CG repeat simulations, each with different velocities, where performed using version 2.1 of the MARTINI model for protein and lipids.\textsuperscript{37,38,42} To maintain a stable tertiary structure of the SERCA CG model, an ELNEDIN model\textsuperscript{40} was used with the elastic network scaffold parametrized with a force constant of $K_S = 1000$ kJ·mol$^{-1}$·nm$^{-2}$ and the cut-off $R_C = 0.7$ nm.\textsuperscript{145} The mapping of cholesterol\textsuperscript{150} and the lipid molecule used for the simulations\textsuperscript{42} is shown in Figure 4.3. The CG beads are given a type based on their physiochemical properties, which determines their interaction strength.\textsuperscript{37,38,42,150} The protein is described similarly, as it is done for protein model.\textsuperscript{38} For MARTINI simulations it has been shown that the simulation time typically should be multiplied by a factor of four to roughly account for the increase in diffusion observed for CG water beads.\textsuperscript{42} The time and figures used in the analysis throughout the work presented in this chapter will therefore be referred to as the “real” time. The simulations were carried out using the GROMACS v. 4.5.5 simulating package.\textsuperscript{72,73}

#### 4.2.1 Protocol

As one of the aims of this study is to uncover if cholesterol could take the place of thapsigargin, simulations are set up with a SERCA structure co-crystallized with a thapsigargin, (and thus trapped in the so-called “E2” state) (PDB: 2CSK\textsuperscript{116}). The initial setup, beside SERCA in “E2” form, also included 370 POPC lipid molecules and 37 cholesterol molecules spread out in the simulation box with random positions and orientations. Counter ions (12 Na$^+$) were added to neutralize the system and it was solvated with standard CG water beads, each representing four water molecules.\textsuperscript{37,42} Following, a short minimization, 30 ns (120 ns of “real” time) of membrane equilibration was carried out with position restraints ($1000$ kJ·mol$^{-1}$·nm$^{-2}$) applied to all protein beads, using
4.2. METHODS

15 fs time step. The simulation was then extended for another 300 ns (1.2 µs of “real” time), with the same time step, and without any position restraints on the protein.

4.2.2 Simulation Parameters

Non-bonded interactions were cut off at 1.2 nm and shifted from 0.9 nm for the Lennard-Jones potential and from 0.0 nm for the electrostatic potential, with a relative dielectric constant of 15. The neighbor list was updated every 10 steps. The temperature was kept constant at 323 K. The pressure during self assembly was kept set to isotropic and kept constant at 1 bar. After bilayer formation, pressure coupling was changed to semiisotropic with the same constant value. Both, temperature and pressure were control by Berendsen coupling algorithm with time constants $\tau_T = 1$ ps for temperature and $\tau_p = 5$ ps for pressure.

4.2.3 Analysis

For each of the 30 self-assembly simulations, 101 snapshots evenly spread over the last 0.6 µs of simulation were used for analysis of SERCA-cholesterol interaction. For each snapshot, data regarding protein residues in contact with cholesterol were collected. To capture the specificity in the cholesterol interaction, individual data were collected for three different parts of cholesterol, namely the polar head (represented by ROH bead type), the ring system (represented by five beads of type R1-R5), and the hydrophobic tail (represented by two beads of type C1-C2) (see Figure 4.3B). If any protein bead was within 6 Å of a cholesterol bead, it was counted as an interaction.
4.3 Results and Discussion

4.3.1 Bilayer Self-Assembly

Previous CG studies of a lipid bilayer self-assembly around membrane proteins have been successful in predicting the accurate position of protein in the bilayer. This approach was applied for the self-assembly of POPC-cholesterol bilayer around SERCA (Figure 4.4). To assure the correct direction of self-assembly of lipid bilayer around protein, protein is kept in the middle of the box, oriented along the z-axis, using position restraints on all protein beads. The formation of bilayer begins when randomly positioned membrane components starts to assemble around SERCA within the first few ns of a simulation. Around 20 ns, a preliminary bilayer has formed and at the 120 ns a full bilayer is formed. To further relax the lipid bilayer and to allow equilibration of lipid-protein interactions, each of 30 simulations was extend for another 1.2 μs.

![Figure 4.4 - Snapshots from the simulations of the POPC:cholesterol bilayer self-assembly around SERCA. SERCA is shown as backbone trace and colored by domains as follows: A-domain is shown in yellow, N-domain in red, P-domain in blue, TM1-2 in pink, TM3-4 in green, TM5-6 in violet and TM7-10 in light brown. Lipids are shown in cyan sticks representation with the phosphate bead shown as a purple sphere. Cholesterol molecules are shown in blue sticks representation with the polar head bead shown as a red sphere. In the initial setup (0 ns) SERCA is positioned in the middle of the box and lipids and cholesterol molecules are scattered around with random position and rotation. During following ns POPC and cholesterol molecules starts to form bilayer around SERCA. The snapshot at 120 ns of simulation shows final step of bilayer self-assembly, just before the position restraints on the protein backbone beads are released.](image-url)
4.3. RESULTS AND DISCUSSION

4.3.2 Transmembrane Surface of SERCA

The TM surface of SERCA has a number of cavities or pockets, which could provide binding sites that would accommodate cholesterol molecules more readily than lipids. One of these sites is the thapsigargin binding pocket. Thapsigargin binds with sub-nanomolar affinity and stabilizes the enzyme in a conformation incompatible with calcium binding, thus effectively inhibiting the protein function. Studying the TM domain of SERCA structure in “E2” form (PDB: 2C8K), three cavities apart from the thapsigargin pocket can be seen in the simulations. All four cavities are depicted in Figure 4.5. The packing of α-helices TM8-10 shows a cleft between TM9 and TM10 with TM8 forming the back of the cleft. Furthermore, the loop between TM8 and TM9 is positioned in the membrane-water interface on the cytoplasmic side of the membrane, forming a roof for the cleft which hinders lipids from effectively filling it up. This cleft is seen in Figure 4.5A, and will be referred to as the C-terminal pocket. The site, formed by TM2, TM4, TM6 and TM9 (Figure 4.5B), has for long been suspected to be the binding site for the transmembrane polypeptides sarcolipin and phospholamban, known to be regulators of SERCA activity. This could mean that other molecules interacting in this site could affect the regulation of SERCA. The last site formed by the tilting of TM1 and TM3 forms a small cavity in the middle of the bilayer, with TM4 forming the back of the cavity (Figure 4.5C).

![Figure 4.5](image)

These pockets change slightly in other conformational states of SERCA, as a result of the structural changes. Most noticeable is a more narrow sarcolipin pocket in the calcium bound so-called “E1” form (PDB: 1T5S). Also, due to up-down translation of helices, the composition of the pocket formed by TM1 and TM3 is different in the conformation where SERCA is phosphorylated and open to the luminal side (“E2P” form, PDB: 3B9B).

4.3.3 Possible Interaction Spots for Cholesterol

To produce a proper sampling of SERCA-cholesterol interaction, 30 independent simulations (resulting in 39.6 µs of simulation in total) were carried out. For the last 0.6 µs of the simulations (18 µs in total), data were collected on specific SERCA-cholesterol
interactions as described in Subsection 4.2.3. For the interactions that last for at least 30% of the simulation the contacts between protein residue and cholesterol specific parts i.e. head group, ring system, and tail, are listed in Table 4.1. The most frequent interactions were found between the ring system of cholesterol and the side chains of Leu913, Val950, Pro952, Leu953, Leu975, and Ile978, with interactions observed in 50-53% of the analyzed snapshots.

Table 4.1 - Number of registered contact events between protein and cholesterol molecules in the 3030 analyzed snapshots

<table>
<thead>
<tr>
<th>Protein Residue</th>
<th>Cholesterol Parts</th>
<th>Polarn ROH (ROH)</th>
<th>Ring (R1-R2)</th>
<th>Tail (C1-C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg63</td>
<td></td>
<td>1119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile64</td>
<td></td>
<td></td>
<td>1014</td>
<td></td>
</tr>
<tr>
<td>Leu66</td>
<td></td>
<td></td>
<td>1063</td>
<td></td>
</tr>
<tr>
<td>Leu67</td>
<td></td>
<td></td>
<td>1319</td>
<td></td>
</tr>
<tr>
<td>Cys70</td>
<td></td>
<td></td>
<td>944</td>
<td></td>
</tr>
<tr>
<td>Phe256</td>
<td></td>
<td></td>
<td>1257</td>
<td></td>
</tr>
<tr>
<td>Lys262</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser265</td>
<td></td>
<td></td>
<td>1484</td>
<td></td>
</tr>
<tr>
<td>Leu266</td>
<td></td>
<td></td>
<td>1302</td>
<td></td>
</tr>
<tr>
<td>Val933</td>
<td></td>
<td></td>
<td></td>
<td>964</td>
</tr>
<tr>
<td>Leu96</td>
<td></td>
<td></td>
<td></td>
<td>1177</td>
</tr>
<tr>
<td>Ala313</td>
<td></td>
<td>1044</td>
<td></td>
<td>1147</td>
</tr>
<tr>
<td>Thr317</td>
<td></td>
<td></td>
<td></td>
<td>1185</td>
</tr>
<tr>
<td>Trp794</td>
<td></td>
<td></td>
<td></td>
<td>1062</td>
</tr>
<tr>
<td>Leu797</td>
<td></td>
<td></td>
<td>1115</td>
<td>933</td>
</tr>
<tr>
<td>Val798</td>
<td></td>
<td></td>
<td></td>
<td>1350</td>
</tr>
<tr>
<td>Leu802</td>
<td></td>
<td></td>
<td>1199</td>
<td></td>
</tr>
<tr>
<td>Thr805</td>
<td></td>
<td></td>
<td></td>
<td>910</td>
</tr>
<tr>
<td>Val990</td>
<td></td>
<td>1074</td>
<td></td>
<td>1559</td>
</tr>
<tr>
<td>Asp951</td>
<td></td>
<td>928</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro952</td>
<td></td>
<td>1165</td>
<td></td>
<td>1608</td>
</tr>
<tr>
<td>Leu953</td>
<td></td>
<td></td>
<td></td>
<td>1576</td>
</tr>
<tr>
<td>Cys910</td>
<td></td>
<td></td>
<td></td>
<td>1342</td>
</tr>
<tr>
<td>Leu913</td>
<td></td>
<td></td>
<td></td>
<td>1625</td>
</tr>
<tr>
<td>Asn914</td>
<td></td>
<td>1280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu922</td>
<td></td>
<td>1277</td>
<td></td>
<td>1303</td>
</tr>
<tr>
<td>Met923</td>
<td></td>
<td>1389</td>
<td></td>
<td>931</td>
</tr>
<tr>
<td>Pro926</td>
<td></td>
<td></td>
<td></td>
<td>1101</td>
</tr>
<tr>
<td>Pro927</td>
<td></td>
<td></td>
<td></td>
<td>1344</td>
</tr>
<tr>
<td>Leu934</td>
<td></td>
<td></td>
<td></td>
<td>1346</td>
</tr>
<tr>
<td>Leu946</td>
<td></td>
<td></td>
<td></td>
<td>1111</td>
</tr>
<tr>
<td>Leu975</td>
<td></td>
<td></td>
<td></td>
<td>1545</td>
</tr>
<tr>
<td>Pro976</td>
<td></td>
<td></td>
<td></td>
<td>1138</td>
</tr>
<tr>
<td>Ile978</td>
<td></td>
<td>1333</td>
<td></td>
<td>1614</td>
</tr>
<tr>
<td>Glu982</td>
<td></td>
<td>1472</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Underlined values indicate the most frequent interactions that were found between the ring system of cholesterol molecules and protein side chain beads, with interactions observed in 50-53% of the analyzed snapshots. Different colors indicate different binding pockets. Light green indicates the pocket between TM1 and TM3, light brown indicates the sarcolipin pocket, and light violet indicates the C-terminal binding pocket.

These six residues constitute two hot-spots of interaction, as three of the residues are found in the loop between TM9 and TM10 in the bottom of the sarcolipin pocket; Val950,
Pro952 and Leu953, whereas the other three are found in the upper part of the C-terminal pocket. In Figure 4.6 the atomistic crystal structure (PDB: 2CSK\textsuperscript{116}) has been colored to reflect the level of specific cholesterol interaction for each residue observed in the CG simulations. From Figure 4.6A it can be seen that the cavity between TM9 and TM10 shows specific cholesterol interaction especially in the upper part of the pocket, and this is also where one of the interaction hot-spots is found. In Figure 4.6B, the surface rendering of the protein makes it clear that the sarcolipin pocket is wide and stretch almost all the way through the bilayer, as opposed to the other cavities on the transmembrane surface of SERCA. Here the specific cholesterol binding is most pronounced in the “bottom” of the cavity, where the other interaction hot-spot is found, but also other residues in the cavity show high levels of specific cholesterol interaction. In Figure 4.6C it can be seen how residues on both sides of the “top” of the TM1-TM3 cavity show a high level of specific cholesterol interaction, whereas no specific cholesterol interaction is observed in the bottom of the cavity. The most striking, however, is that the thapsigargin binding pocket Figure 4.6D has a very low level of cholesterol interaction, despite the hypothetical good fit of cholesterol in this pocket. 12, 14 and nine residues in or around the C-terminal, sarcolipin and TM1-3 pockets, respectively, show specific interaction with cholesterol in more than 30\% of the analyzed snapshots, whereas none of the residues in or around the thapsigargin binding pocket show cholesterol interaction at this level (see Table 4.1).

If cholesterol was excluded from the protein surface by POPC lipids, as suggested by \textit{e.g.} lipid substitution experiments combined with ATPase activity measures\textsuperscript{156} and tryptophan fluorescence quenching experiments,\textsuperscript{139} only the cavities should show white to red coloring in Figure 4.6, but it appears not to be the case.

### 4.3.4 Annular Interactions

As mentioned before, the lipids in the annulus are indistinguishable from the bulk lipids by NMR techniques. It has been, however, estimated experimentally to be between 18 and 40 sites, depending on the methods applied.\textsuperscript{140,156--158} It is also clear from the simulations that it is not trivial to distinctly define when a lipid is interacting with
the protein and takes part in the lipid annulus around the protein. The membrane components pack tightly around the protein, and as the lipids are very flexible molecules, the surface of the protein is in the simulations in contact with on average 57 lipids and seven cholesterol molecules. As many of these molecules in reality rather belong to the bulk and are only in contact with the protein very briefly and with a very little interaction surface, the definition of protein interaction can be sharpened to demand an interaction interface where at least 35% of the molecule (three beads for cholesterol, five beads for lipids) is in protein contact (less than 7 Å between protein and lipid beads) for at least 12 ns. Likewise, a break of contact for less than 12 ns is not considered dissociation from the protein surface. Using this definition, an average of $32.0 \pm 2.6$ POPC lipids and $6.5 \pm 1.7$ cholesterol molecules compose the lipid annulus around the protein, and an example of a lipid annulus adhering to this definition is shown in Figure 4.7B. However, the number of annular binding sites is sensitive to the annulus definition, and if it is instead required that at least 50% of the molecule should be within 7 Å of the protein before it is considered as participating in the annulus, then the annular composition changes considerably to on average $21.9 \pm 2.6$ lipids, but still $6.1 \pm 1.7$ cholesterol molecules. Still, for both definitions, the obtained number of sites is within the range of experimentally determined values. To illustrate how the components in the annulus is exchanged with bulk molecules on the microsecond time scale described by the simulations, the participation pattern for POPC lipids and cholesterol molecules in the lipid annulus (based on the 35% molecular interaction interface) in one of the simulations is depicted in Figure 4.7A.

![Figure 4.7](image_url)

**Figure 4.7** - Membrane components interacting with SERCA. (A) Graph of interaction lifespan for lipids (blue) and cholesterol (red) having close interaction with SERCA (as defined in the text) in the last 0.6 µs of one of the CG self-assembly simulations. A line is present as long as the molecule interacts with SERCA. (B) and (C) Snapshot of the membrane molecules in close interaction with SERCA defining an annulus around the protein. (B) The protein shown in gray surface representation with only the transmembrane part shown. Lipids represented in cyan sticks with the phosphate beads as purple spheres. Cholesterol molecules shown in blue sticks with the polar head as red spheres. (C) The same snapshot as in B, but viewed along the membrane normal and omitting the protein.

It can be seen that initially 20 lipids and six cholesterol molecules participate in the annulus, and while 14 different cholesterol molecules take part in the annulus during the
0.6 μs, in the end five of the initially interacting molecules have kept or re-established their annular binding, and only one seems to have left completely and is replaced by another. More than 110 different lipid molecules participate in the annulus during the 0.6 μs, and only two of the lipids from the initial annulus stay with the protein all through the 0.6 μs.

To see how the cholesterol molecules affect the annulus composition, a pure POPC bilayer was allowed to self-assemble around SERCA, in a setup identical to the POPC-cholesterol CG MD simulations, except the cholesterol molecules were replaced by POPC lipids. For this simulation, the lipid annulus contained on average 34 lipids if at least 35% of the lipid should show protein contact, and on average 25 lipids if at least 50% molecular contact was required. Comparing the number of annular binding sites obtained for bilayers with and without cholesterol, it is evident that either the cholesterol molecules replace POPC lipids in the annulus in a ≈2.5 : 1 manner, rather than in a 1 : 1 manner, as previously assumed in the modelling of experimental data,139,157 or a number of cholesterol molecules find binding spots not accessible to lipids. Based on visual inspection of how POPC lipids and cholesterol molecules interact with the protein, it does not seem likely that they would replace each other in a 1 : 1 manner. However, it also seems plausible that a cholesterol molecule could snug in places on the rugged protein surface, which would not interfere too much with lipids participating in the annulus by the definition used here. Based on the simulations, it thus seems to be a combination. It should also be emphasized that the positions of the membrane components in the annulus is not a static picture with fixed binding sites, and many different presentations of the lipid annulus like the one in Figure 4.7B can be produced, not one being more correct than the other.

4.3.5 Potential Non-annular Binding Interactions

The analysis of the number of annular binding sites given above did not distinguish between annular and non-annular binders, as both would show the high level of protein interaction required by the annulus definition. This, however, will also be the case for the experimental methods used to estimate the number of binding sites, as they also rely on quantifying the amount of lipid in contact with the protein surface. The annulus participation pattern given in Figure 4.7A would therefore also include potential non-annular binders. The SERCA-cholesterol interactions interpreted as non-annular binding in tryptophan fluorescence quenching experiments correspond to on-times longer than the exchange time for phospholipids between the bulk phase and annular sites on SERCA, that is ≈ 0.1 μs.139 However, the CG simulations presented here do not show a clear separation between groups of annular and non-annular binding events (data not shown), the more significant cholesterol interactions should be found among the binding modes observed to be stable on the analyzed time scale of 0.6 μs.

To extract candidates for cholesterol specific non-annular binding sites, the 57 cholesterol and 50 lipid molecules registered to participate in the annulus, where at least 35% of the
POPC or cholesterol molecule is at most 7 Å from protein for at least 12 ns, all through the last 0.6 µs of simulation were inspected visually. In each case it allowed to determine if the long-time protein interaction was due to diffusion on the surface of the protein or due to a stable binding mode. 18 of the inspected cholesterol molecules and 19 of POPC lipids were judged to be potential non-annular binders, while the rest of the cases were surfing the protein surface and visiting different binding modes during the 0.6 µs. Snapshots of the potential non-annular binders are collected in Figure 4.8 - Figure 4.10. The thapsigargin binding pocket is in 12 of the 30 simulations occupied by a lipid, all through the last 0.6 µs of the simulation, with a characteristic binding mode immobilizing one of the lipid tails in the cavity (see Figure 4.8), while the other tail is flexible, and with interactions between the phosphate bead of the lipid and Ser830 or Lys252. This binding mode has been previously observed in one “E2” crystal structure stabilized by cyclopiazonic acid (CPA) (PDB: 2EAU), where non-annular lipid is present in the free thapsigargin binding pocket, but with the head-group more deeply buried than seen in the simulations. Only in one case a cholesterol molecule is stuck in the cavity for the entire 0.6 µs (Figure 4.8), and it seems clear from these observations together with the general low frequency of cholesterol interactions illustrated in Figure 4.6, that this area of the protein surface has a higher relative affinity towards the POPC lipids compared to the cholesterol molecules.
4.3. RESULTS AND DISCUSSION

The pocket found between TM1 and TM3 has in six of the simulations a cholesterol molecule more or less inserted throughout the 0.6 µs (Figure 4.9A). Three of these cases share binding modes, where the polar head of cholesterol is squeezed in between the helices, interacting with the backbone beads of Ala303 on TM4, Ser261 and Ile264 on TM3 and Ile64 on TM1. In five of the simulations the end of a lipid tail is instead trapped in this cavity, while the rest of the lipid is free to move. In the C-terminal pocket only two cholesterol molecules were found to interact with TM surface of SERCA throughout the 0.6 µs (Figure 4.9B), and they are positioned with the long axis of the molecule almost perpendicular to the membrane normal.

![Figure 4.9 – Potential non-annular binders found in (A) pocket between TM1 and TM3 and (B) C-terminal binding pocket. The coloring of the molecules and the time frame are as described in Figure 4.8.](image)

The sarcolipin binding pocket, is a large cavity, where several cholesterol molecules could fit at the same time, and in fact this is also frequently observed (Figure 4.10). Moreover in four of eight cases cholesterol molecules share almost identical binding mode. The four other cases show distinctly different positions of cholesterol molecules, one of which has the polar head stuck between TM2 and the luminal loop of TM9-TM10, interacting with one of the cholesterol interaction hot spots (Figure 4.6B). A high cholesterol interaction frequency, as is observed for the three residues in the luminal TM9-TM10 loop, does not distinguish between frequent short visits by cholesterol or long encounters with the same cholesterol. For the cholesterol interaction hot-spot found in the bottom of the sarcolipin binding pocket, the interaction seems mainly to arise from cholesterol molecules loosely bound or passing by. In many of the cases where cholesterol showed protein interaction all through the 0.6 µs, but was not deemed to be in a stable binding mode, the molecule was instead moving around in the sarcolipin pocket, occupying various positions and
orientations with respect to the membrane. For the top part of sarcolipin pocket, found between TM9 and TM10, no lipids are seen to stick all through the 0.6 µs. This is the second hot-spot for cholesterol interaction, and in two cases cholesterol molecules are resting in this pocket in a characteristic binding mode, with the long axis of the molecule almost perpendicular to the membrane normal.

![Figure 4.10: Potential non-annular binders found in sarcolipin pocket. The coloring of the molecules and the time frame are as described in Figure 4.8.](image)

It seems there is a cavity where cholesterol can fit in nicely, as opposed to the lipids, which are forced to have their ionic head-group in the interface to water, and therefore seldom dislocates vertically to a great extend. For the cholesterol analogue, cholesterol hemisuccinate, the number of cholesterol specific non-annular binding sites on the surface of SERCA has been estimated to between one and seven, and an estimate based on the above described observations would be within the same range. In only two out of the 30 simulations, lipids are observed to be caught in this pocket, one time with the head-group between TM2 and TM4, and the other instance between TM4 and TM9.

To relate the observations from a series of tryptophan fluorescence quenching experiments described in all Section 4.1.1, it is relevant to consider the position of the tryptophan residues with respect to the binding pockets. Figure 4.11 shows the binding pockets with a 10 Å radius sketched around all tryptophan residues found in the vicinity of the pocket. The efficiency of tryptophan fluorescence quenching depends on the distance between tryptophan and the dibromo group to the sixth power, with a quenching efficiency of 50% at a distance of 8 Å. As the quenching is expected to comply with Förster energy transfer theory, the quenching efficiency would be less than 20% for distances larger than 10 Å and almost gone when the distance to the dibromo group is more than 14 Å. Based on the potential non-annular binders observed in the simulations, the SERCA-cholesterol interaction resulting in an increase in quenching efficiency when adding brominated sterol to a brominated lipid bilayer can now be deduced. The potential non-annular binding modes observed in the thapsigargin binding pocket and the pocket between TM1 and TM3, all seem to have lipid counterparts, which could
also fill out the pocket, and an increased quenching by the addition of brominated sterol to brominated lipid could thus not be expected. Furthermore, the thapsigargin binding pocket has no tryptophan residues near enough for a binding to result in significant quenching, as seen in Figure 4.11. However, the potential non-annular binding modes observed in the sarcolipin binding pocket and the C-terminal pocket would have a good chance of affecting the tryptophan fluorescence, and both these pockets show a higher preference towards immobilizing cholesterol molecules than POPC lipids.

Figure 4.11 - The surface of the transmembrane part of the protein is rendered in the same manner and view angle as in Figure 4.12, and the tryptophan residues are shown in VdW rendering with carbon atoms in cyan, oxygen atoms in red, and nitrogen atoms in blue. The tryptophan residues are surrounded by a transparent sphere with a radius of 10 Å. (A) Thapsigargin binding pocket, (B) Binding pocket between TM1 and TM3, (C) sarcolipin binding pocket, (D) C-terminal binding pocket.

4.4 Perspectives

For the three binding pockets showing the highest level of cholesterol interaction in the CG simulations, the representatives of potential non-annular cholesterol molecules (described in Subsection 4.3.5) shown in Figure 4.12 were selected for further studies of SERCA-cholesterol interaction in atomic detail. As the study concerns only the protein interaction with this particular cholesterol molecule, the entire CG setup was not reversed from the CG description to a corresponding setup with atomistic description, instead, for each of the three snapshots, the atomistic protein structure used for the CG setups (PDB: 2C8K) was aligned to the CG protein structure with respect to the Cα’s of the backbone beads in the transmembrane part of the protein. Correspondingly, an atomistic cholesterol structure was aligned to the CG cholesterol interacting with SERCA in the snapshot. The position of the protein with respect to the membrane observed in the CG simulation was maintained in the atomistic setup, as the membrane normal was kept, and the phosphate bead positions were used to direct the position of the generated atomistic POPC bilayer.
The atomistic studies of SERCA-cholesterol interaction are undertaken by Ph.D. student Henriette Autzen, and the details are thus not included in this thesis.

4.5 Conclusions

The CG MARTINI model allowed for simulations of the self-assembly of a bilayer composed of POPC and cholesterol molecules around SERCA. The results show, that based on the definition of annular sites used in this work, an average of 32.0 ± 2.6 POPC lipids and 6.5 ± 1.7 cholesterol molecules compose the lipid annulus. Furthermore, several cases of both lipids and cholesterol molecules were judged to interact especially with the protein as potential non-annular binders.

Nothing in our simulations suggests that the probability of finding cholesterol in the lipid annulus is lower than finding cholesterol anywhere else in the bilayer, rather the on-times for cholesterol molecules are consistently longer than those observed for the POPC lipids. The longer on-times, of course, fit with the experimental interpretation of cholesterol interaction being of a non-annular character. The main reason why the experimental binding models fit a relative binding constant for sterol over DOPC of around 0.6 or less\textsuperscript{127,157} could be the underlying assumption that a cholesterol would be able to replace a lipid on the surface of the protein in a one-to-one manner, while in reality this is not likely to be the case, as both the structure and physical properties of a phospholipid and cholesterol are fundamentally different.

Further investigation of SERCA-cholesterol interactions in atomic detail are in progress and will help define the exact binding modes of cholesterol, and the potential impact of cholesterol binding on the protein structure and dynamics.
Chapter 5

Dynamics of Saposin-Like Proteins

The work presented in this chapter has been performed during my stay in Professor Peter Telemann’s group at the University of Calgary, and concerns the dynamics of saposin-like proteins.

Contents

5.1 Saposins A, B, C, D, and NK-lysin .......................... 70
  5.1.1 Hypothetical Binding Modes to Biological Membranes ... 72
5.2 Methods ...................................................... 74
  5.2.1 System Setups ............................................ 74
  5.2.2 Simulation Protocol ...................................... 74
  5.2.3 Analysis ................................................... 75
  5.2.4 Abbreviations ............................................. 75
5.3 Structure Changes in AMBERSB99-ILDN and GROMOS45a3 Force Fields .............................................. 76
5.4 MARTINI CG Simulations ..................................... 81
  5.4.1 Water Model Influence on Protein Structure .............. 83
  5.4.2 Biological Implications ................................... 83
5.5 Conclusions ................................................... 84
5.1 Saposins A, B, C, D, and NK-lysin

Sphingolipid activator proteins (saposins) A, B, C, and D,\textsuperscript{162-165} are generated from a common precursor, prosaposin, in acidic organelles such as late endosomes and lysosomes.\textsuperscript{162,166,167} They are small \(\approx 80\) amino acid long, water soluble, non-enzymatic proteins, highly homologous with up to \(\approx 60\%\) amino acid similarity of the primary sequences.\textsuperscript{168} Along with the tumorylitic proteins \textit{e.g.} NK-lysin protein,\textsuperscript{81} and regulators of membrane surface activity, \textit{e.g.} surfactant-associated protein B (SP-B),\textsuperscript{82} they belong to a large and diverse family of saposin-like proteins that share both, an ability to interact with biological membranes,\textsuperscript{163,169} and similarities in their three-dimensional structure. The polypeptide fold is composed of four to five amphipathic \(\alpha\)-helices forming a central hydrophobic cavity, and a set of conserved cysteine residues that form intramolecular disulfide bridges (Figure 5.1). This \(\alpha\)-helical bundle and set of cysteine residues, referred to as the “saposin fold”;\textsuperscript{81} is common for all the saposin-like proteins and enables them to interact with membranes.\textsuperscript{170} However, saposin-like proteins exhibit a wide variety of biological functions, resulting from less conserved, solvent exposed side chains, and their interaction with membranes depends on the lipid composition, lipid/protein ratio, pH and buffer composition (for reviews see: ref.\textsuperscript{163,164}). For instance saposin A is known to facilitate the degradation of \(\beta\)-glucosyleramidase and \(\beta\)-galactosyleramidase,\textsuperscript{171} while saposin B stimulates the breakdown of sulfatide by arylsulfatase A, globotriaosylceramide and galabiosylceramide by \(\alpha\)-galactosidase A, and lactosylceramide by galactosyleramidase \(\beta\)-galactosidase.\textsuperscript{172,173} In general it is believed that in the absence of saposins, the oligosaccharide chains of the membrane-bound lipids do not extend far enough into the lysosomal lumen to be accessible to the active sites of the hydrolases. This leads to lysosomal storage disorders like Krabbe\textsuperscript{174} or Farber\textsuperscript{175} disease, which are caused by a deficiency of saposins A and D, respectively.
Figure 5.1 — Sequence and structures of Sap A-D and NK-lysin. (A) Multiple sequence alignment of four human saposins and porcine NK-lysin using ClustalW. The alignment shows conservation of six cysteine residues marked in light gray with indicated connectivity and named C-C1, C-C2, and C-C3, for disulfide bridges one, two, and three, respectively. Based on X-ray structures of Sap A-D and a NMR structure of NK-lysin the sequences corresponding to helices H1, H2, H3, and H4 are shaded blue, green, yellow, and red, respectively. For NK-lysin the helix H3 is composed of two helices spanning residues 42-51 and 53-61 (as reported in NMR structure file), however for simplicity we will consider it as one helix. The dark gray regions in saposin A, B and D indicate a turn of $3_{10}$ helix present at the end of helix H2 in saposin A, B, and D, and beginning of helix H3 in Sap D. The pink color indicates the conserved tyrosine 54 that is found at the position of the kink in H3 of Sap A, C, and D and in Sap B chain A, and at phenylalanine 54 in NK-lysin. The lower bars indicates stem (red) composed of H1 and H4 and a hairpin (black) composed of H2 and H3. (B) NewCartoon representation of saposin A, B (CC’ dimer), C, D, and NK-lysin with helices colored according to scheme in A.
CHAPTER 5. DYNAMICS OF SAPOSIN-LIKE PROTEINS

5.1.1 Hypothetical Binding Modes to Biological Membranes

In X-ray and NMR structures of saposins and NK-lysin, the proteins are found in a variety of conformations, ranging from a monomeric, substrate-free closed or open structure to a dimeric, V-shaped, ligand-bound conformation, with different degrees of opening of the hydrophobic cavity. Comparison of the different conformations suggests that the conformational changes involved in the opening and closing of the protein are facilitated by the flexibility of the hinge region (the loops between helices H1/H2 and H3/H4) and the kinking of H3 helix, at conserved Tyr54 and Phe54 in saposins and NK-lysin (Figure 5.1), respectively. The transition between open/closed conformations appears to be restrained only by the disulfide bridges (Figure 5.1).

Different binding modes for binding to biological membranes were proposed for saposins A, B, C, D and NK-lysin, suggesting that the orientation of helices with respect to the biological membrane influence the biological function of protein. Biophysical studies of saposin D, for instance, documented that saposin D acting as a "solubilizer" can break down phospholipids-containing membranes at critical phospholipid concentrations. The solubilization of membranes was shown to be effective only at low pH and in the presence of anionic lipids. Based on the existing studies of saposin D, a mechanism of interaction for saposin D with membranes has been proposed, as is shown in Figure 5.2A. The low lysosome pH neutralizes the negative charge on glutamate residues, inhibiting the repulsion of saposin D by anionic membrane. Prior to membrane binding saposin D is found in a monomer-dimer equilibrium. The interaction starts when the proposed saposin D dimer, with its positively charged "bottom", approaches the surface of the intra-lysosomal membranes, which are rich in negatively charges. In the next step the dimer rotates 180° around its long axis on the membrane surface, such that its "top", which is rich in apolar residues, is buried in the membrane bilayer and the positively charged residues are exposed to the solvent. The dimer's interactions with the membrane lead to a weakening of the monomer-monomer interactions, which results in structural rearrangements of the saposin D dimer. Each monomer of the saposin D dimer opens, and dips the hydrophobic surfaces of the α-helices into the membrane, thereby perturbing the membrane structure. Finally, the saposin D monomer changes configuration to the closed form, lifts a lipid out of the membrane, and leaves the membrane with a bound lipid.

In contrast to saposin D, saposin C does not act as a lipid solubilizer, but rather participates in the fusion of vesicles containing anionic phospholipids. The "clip-on" model (Figure 5.2B) proposed for saposin C, based on the fluorescence analysis, suggests that two saposin C molecules are able to destabilize and fuse anionic vesicles at lysosomal pH. The initial binding of saposin C to negatively charged membranes occur when two monomers of saposin C rapidly anchor with their helices H1 and H4 (with the nomenclature given in Figure 5.1) into two different phospholipid vesicles. Next, the two saposin C molecules clip to each other by interactions between the helices H2 and H3, bringing the vesicles close enough for a fusion.
5.1. SAPOSINS A, B, C, D, AND NK-LYSIN

Figure 5.2 – Schematic model of the proposed mechanism for interactions between membranes and Sap D and Sap C, respectively. (A) Lipid perturbation of a membrane by Sap D. (1) Water-soluble Sap D monomers and dimers bind to a negatively charged membrane surface, and (2) rotate so that the hydrophobic “top” of the dimer faces the membrane surface. (3) Sap D changes configuration to a boomerang shape, and amphipathic α-helices stretch parallel to the lipid bilayer, exposing polar residues to the solvent. The hydrophobic surface dips into the membrane and perturbs its structure. (4) Finally, Sap D changes conformation to the closed form, lifts a lipid out of the membrane, and leaves the membrane with a bound lipid. (B) “Clip-on” model for Sap C-induced vesicle fusion proposed in 2003 by Wang et al.. Sap C molecules anchored to phospholipid bilayers of vesicles clip to each other through domain swapping, bringing the vesicles close enough to fuse. The size of the vesicle and saposins are not to scale. Reprinted from Rossmann et al., with permission from Elsevier.

This is not the only example of fusogenic behaviour of saposin-like proteins. Similar behaviour which has been shown to fuse liposomes containing DPPC and POPC lipids is known for SP-B. MD studies of hemifusion induced by SP-B proteins had led to a proposed mechanism of action. The fusion of vesicles occurs when the vesicle comes closed to solvent exposed helices H2 and H3 of a closed conformation of SP-B that is already anchored to another vesicle with the C- and N-termini (helices H1 and H4, respectively). The SP-B inserts into the headgroup region of the vesicle, rearranging as a wedge between the two vesicles, and causing formation of a lipid bridge connecting the outer leaflets of the vesicles. During the lipid flow, these leaflets merge to form a hemifusion stalk that is stabilized by SP-B which changes conformation from closed to open. The vesicles fuse to finally form an elongated stalk, with the protein lining a positively curved side of the stalk.

The mechanism of membrane binding and lipid extraction for saposin-like proteins is not fully understood. It is believed that conformational changes leading to an opening of a hydrophobic cavity are important for the recognition and extraction of lipids from mem-
branes, not only for saposins, but for all saposin-like proteins as they share structural similarities. Both the flexibility of existing conformations of saposins and NK-lysin, as well as the proposed models of binding to the biological membranes seems to support this belief. However, the dynamics of the proteins is still difficult to study by experimental methods, therefore in this project the dynamics of saposin-like proteins, namely saposins A, C, D, and NK-lysin, in monomeric closed conformation, using molecular dynamics simulations, at both atomistic and CG level is studied. The X-ray and NMR structures of saposin-like proteins in combination with MD studies, can be used to better understand the dynamics and function of proteins such as SP-B, for which the high-resolution structure is still unknown.

5.2 Methods

5.2.1 System Setups

Four different setups were built starting from the closed, monomeric, unliganded X-ray structures of human saposin A (PDB: 2DOB, chain A), C (PDB: 2GTG, chain A), D (PDB: 3BQP, chain A), and NMR structure of porcine NK-lysin (PDB: 1NKL, conformer #1) (Figure 5.1). For all four proteins, ionization states of titrable side chains were in their standard sites for pH 7, and the six cysteine residues were connected to form three disulfide bridges. For each of saposin-like protein chain A and water molecules present in the crystal structure were used for setting up atomistic simulations. The CG setups did not include the crystal water molecules. For NK-lysin the first of twenty NMR structures was used as a starting point for both the atomistic and the CG simulations.

5.2.2 Simulation Protocol

MD simulations were performed with GROMACS 4.5.4. in periodic boundary conditions and in explicit water. Four different force fields were used; two atomistic force fields: GROMOS45a3 and AMBER99SB-ILDN, and two CG force field: MARTINI force field version 2.1 and version 2.2. Each saposin like protein was placed in a truncated octahedral box with a minimum distance of 0.8 nm between solute and the box edge. The box was solvated with a force field specific water model and 0.15 M of NaCl, in addition to counter ions to neutralize the system, were added. An equilibration procedure was carried out followed by a short run in which position restraints (1000 kJ·mol$^{-1}$·nm$^{-2}$) were applied to all heavy atoms of protein in atomistic simulations for 50 ps or to all beads of protein in CG simulations for 1 ns. The setups were then simulated with three copies of each (with different initial velocities) for 2 µs with a 2 fs time step for atomistic simulations, and 100 µs with a 25 fs time step for CG simulations (without any restraints).
5.2. METHODS

AMBER99SB-ILDN Simulations

The TIP3P water model was used. The temperature and isotropic pressure were kept constant at 310 K and 1 bar, respectively using the velocity rescaling thermostat with time constant $\tau_t = 0.1$ ps for the temperature and Berendsen coupling algorithm with time constant $\tau_p = 0.5$ ps for pressure. A cut-off of 1.0 nm was used for the Lennard-Jones interaction and short-range electrostatic interactions. The neighbor list was updated every 10 steps. The long-range electrostatics interactions were modeled using PME. Bond lengths were constrained using the LINCS algorithm for the protein.

GROMOS45a3 Simulations

The SPC water model was used. A twin-range cut-off scheme with a van der Waals cut-off radius of 1.4 nm, and a real-space cut-off of 0.8 nm for long-range electrostatics interactions modeled with PME was used. The temperature and pressure were treated as in AMBER99SB-ILDN simulations. The neighbor list was updated every 10 steps.

MARTINI Simulations

Non-bonded interactions were cut off at 1.2 nm and shifted from 0.9 nm for the Lennard-Jones potential and from 0.0 nm for the electrostatic potential, with a relative dielectric constant of $\epsilon_{rel} = 15$ for the standard water model and $\epsilon_{rel} = 2.5$ for the polarizable water model. The neighbor list was updated every 10 steps. The temperature and isotropic pressure were kept constant at 310 K and 1 bar, respectively, using the Berendsen coupling algorithm with time constants $\tau_t = 1$ ps for temperature and $\tau_p = 5$ ps for pressure.

5.2.3 Analysis

A measure of conformational stability of the saposin-like proteins was provided by DSSP analysis of the secondary structure for the whole protein as a function of time for atomistic simulations. The RMSD was measured for N atoms for atomistic and backbone beads for MARTINI simulations, using the whole trajectory. In addition to DSSP and RMSD analysis, a principal component analysis (PCA) on the trajectories was performed using the C atoms or backbone beads for atomistic and MARTINI simulations, respectively.

5.2.4 Abbreviations

Three repeat simulations of each system of saposin A, C, D, and NK-lysin in water for both the atomistic and CG force fields, will be referred to as r1, r2, and r3, respectively. For simplicity the helices H1 and H4 will be referred to as the stem of the protein, and a pair of helices H2 and H3 as the hairpin (Figure 5.1) of the protein. Additionally, the
AMBER99SB-ILDN will be referred to as AMBER, GROMOS45a3 as GROMOS, and two other abbreviations will be used: MS for the MARTINI with standard water model and MP for MARTINI with polarizable water model.

5.3 Structure Changes in AMBERSB99-ILDN and GROMOS45a3 Force Fields

The 2 µs atomistic simulations allowed for detailed investigation of secondary structure changes and the ability of AMBER and GROMOS force fields to reproduce a stable "saposin fold". Several studies of atomistic force fields indicated that the family of AMBER99 force fields strongly favors α-helical structures and prevents the formation of β-sheet structures, while the family of GROMOS force fields favors turn-coil conformations and extended β-sheet structures. The same behaviour was observed in this study. The DSSP analysis of structures in AMBER simulations (Figure 5.3) showed that the secondary structure elements are globally stable, although, some local unfolding of α-helical structures were observed in case of saposin A and C. In the saposin A in r1, r2, and r3 part of helix H3 unfolds, while additionally in r3 the whole helix H4 unfolds. In saposin C helix H1 partially unfolds at the end of r1 simulation. These small changes in secondary structure did not influence the α-helical "saposin fold" which was maintained in all cases. On the contrary the DSSP analysis for GROMOS simulations (Figure 5.4) showed that proteins underwent significant secondary structure changes, mainly from α-helical to β-element structures. In saposin A all three repeat simulations showed formation of β-strands. In the case of r1 the β-strands were formed at the C-terminus and at the loop region between helices H1 and H2. In case of r2 the β-sheet was formed at the loops between helices H1/H2 and H3/H4, and maintained for 0.7 µs of simulation. In the case of r3 a β-strand was formed mainly at the C-terminus. Saposin C showed partial unfolding of helix H1 in r1 and r2 repeat simulations, and local unfolding of helix H3 with the formation of a β-sheet in r3 simulation. In saposin D significant changes to the secondary structure occurred in the r2 simulation where the formation of β-elements at the loop regions between helices H1/H2 and H3/H4 led to the formation of a β-sheet after ≈1.4 µs of simulation. For NK-lysin the overall helical structure was maintained with local unfolding of helix H3 in the r1 simulation, helices H3 and H4 in the r2 simulations, and changes in the α-helical structures from α-helix to π-helix in helices H2 and H3 in the r3 simulation. In all studied proteins in GROMOS force field the formation of β-elements occurred mainly at the loop regions and at the N- and C-termini, which are poorly defined compared to an α-helix. However, the overall helical structure of proteins, was still maintained. The unfolding of helices and secondary structure changes are not repeatable through all three repeat simulations. For instance, for saposin D in the second run coils defined between H1/H2 and H3/H4 helices changed their secondary structure to form a β-sheet, while the two other runs do not present the same changes in the structure. This changes therefore, will be considered as local unfolding.
Figure 5.3 – Secondary structure according to the DSSP algorithm as function of time for AMBERSB99-ILDN simulations, for each of three repeat simulations (r1, r2, r3, respectively, starting from left) of saposin A, C, D, and NK-lys in.
Figure 5.4 - Secondary structure according to the DSSP algorithm as a function of time for GRAMOS45a3 simulations, for each of three repeat simulations.
5.3. STRUCTURE CHANGES IN AMBERSB99-ILDN AND GROMOS45A3 FORCE FIELDS

Circular dichroism (CD) spectroscopic studies have been used to estimate the content of α-helices and β-sheet structures in saposins A, C, D, and NK-lysin, at lysosomal pH\(^\text{180,196,199}\). The CD studies showed that the content of secondary structure elements for NK-lysin is similar with and without presence of lipids, and is as follows: 50-58% α-helix and 24% β-sheet.\(^\text{199}\) One of the first CD studies of saposins by O'Brien \textit{et al.}\(^\text{198}\) revealed that there is 44%, 42%, and 53% α-helix, and 8%, 33%, and 0% β-sheet in the secondary structure of saposins A, C, and D, respectively. Remarkably, more recent studies of CD for saposin A and C by Qi \textit{et al.}\(^\text{180}\) showed significant deviations from these values. The secondary structure of saposin A was shown to be composed of 44% α-helix, 32% β-strand, and 24% of other structure in the absence of lipids. When the same was measured in the presence of lipids the α-helical content dropped to 39%, and the percent values of β-strand and of other structures were increased to 35% and 25.5%, respectively. Additionally 0.5% of β-turn was measured.\(^\text{180}\) Secondary structure of saposin C without lipids was measured to have high content of β-strands (42%) compared to the α-helical content with a level of 30% and other structures at a level of 28%. With the addition of different PS or PC lipids, the content of secondary structure elements of saposin C changes to ≈30-50% α-helix, ≈4-40% β-strand, ≈1-14% β-turn, and ≈28-32% of other structures. Moreover, it was speculated that with increasing pH the helical content of saposins should decrease. This is a bit puzzling, and as there is no structural data at different pH it is difficult to judge which force field is more accurate in this case.

The RMSD analysis (Figure 5.5) reflects the flexibility and stability of proteins in GROMOS and AMBER force fields, respectively. On average, the RMSD values for GROMOS force field are ≈2 Å higher than for AMBER, due to the flexibility of proteins in GROMOS simulations that underwent local unfolding and changes in secondary structure. This was also confirmed by the PCA, which showed that in GROMOS force field the conformational space for the proteins is much bigger than for the AMBER simulations. In the GROMOS simulations all proteins, while in the AMBER simulations only saposin A, as the most flexible one, exhibit twisting motion (Figure 5.6). The twisting motion introduces changes to the tertiary structure of the protein and occurs as follows (Figure 5.6). In the initial conformation the stem formed by helices H1 and H2 is deflected from a parallel position with respect to hairpin formed by helices H2 and H3. During the twisting motion the stem is positioned parallel to hairpin, and then goes back to the initial structure or deflects even more to form a more globular form in which helix H3 bends at the kink position. In this structure the hairpin engulfs the stem and the hairpin loop bends towards the hydrophobic core of the protein.
Figure 5.5 — Running average of RMSD of $C_{\alpha}$-atoms in atomistic simulations (left panel) and BB in coarse grained simulations (right panel) of saposin-like proteins. For each saposin-like protein four plots are presented indicated by letter in upper left corner: A for AMBER99SB-ILDN, B for GROMOS45a3, and C, D for MARTINI with standard and polarizable water model, respectively. Three copies of each simulation are marked as follows: r1 black, r2 red, r3 green. Note that the y-axis runs to 1 $\mu$s for atomistic simulations and 2 $\mu$s for the CG simulations.
5.4 MARTINI CG Simulations

Additionally to the atomistic simulations, 100 µs of MARTINI CG simulations with standard (MS) and polarizable water (MP) models were performed to check the influence of the water model on the tertiary structure changes which are hypothesized to be crucial for membrane binding.

In the MARTINI model the secondary structure is imposed at the beginning of the simulation, and since the secondary structure in the AMBER simulations is stable and only undergoes local changes in the GROMOS simulations it is reasonable to use the MARTINI CG model. The RMSD analysis for the MS simulations showed that the protein in the MS simulations exhibits only twisting motion, as observed in the atomistic simulations (Figure 5.6) with the exception of saposin D which also underwent full opening in r1. In the MP simulations, the proteins showed significant changes in the RMSD values indicating conformational changes, especially for saposin C, D, and NK-lysin (Figure 5.5). These results have been confirmed with PCA analysis (Figure 5.7), which shows that the sampled conformational space is much larger for the MP simulations than for the MS simulations, as proteins underwent also scissor and full opening. The full opening of the protein starts from the initial conformation, which exhibits twisting motions, goes through the twisted semi-open conformation, and ends with a fully open conformation where the hydrophobic core is solvent exposed (Figure 5.7). The main motion in the full opening occurs at the loops between helices H1/H2 and H3/H4 which serves as a hinge for the opening, allowing the stem and hairpin to be pulled outwards relatively to each other. The peaks up to ≈1.1 nm on the RMSD plots for saposins A, C, D, and NK-lysin in the MP simulations (Figure 5.5) indicate scissor-like open conformation of these pro-
teins (Figure 5.7), in which the loops between H1/H2 and H3/H4 move outwards while pulling the attached helices, and the protein opens with the hairpin loop serving as a hinge. The full opening, scissor-like motion, and twisting movement are limited by the three disulfide bridges, C-C1, C-C2, C-C3.

Figure 5.7 – PCA analysis for MARTINI simulations with standard (W) and polarizable (PW) water model. In the left panel (A) the projection of the first eigenvector on the second eigenvector for all backbone beads is shown for each saposin-like protein. Three runs are plotted on the same projection for each of studied proteins, color-coded as follows: black - r1, red - r2, green - r3. The numbers on the projection plots correspond to the cartoon representation of conformational clusters observed during the simulation (B). #1 represents the initial structure of the saposin-like proteins with three disulfide bridges, C-C1, C-C2, and C-C3 (orange). #2 represents scissor opening where the loop between H2 and H3 (blue) serves as a hinge for motion, and red arrows indicate the direction of the opening. #3 represents a twisted semi-open conformation. #4 represents full opening of the saposin-like proteins, in which the stem and the hairpin are pulled in opposite directions as indicated by red arrows. The opening occurs on the hinge region consisting of the loops between helices H1/H2 and H3/H4 (blue).
5.4. MARTINI CG SIMULATIONS

5.4.1 Water Model Influence on Protein Structure

One of the aims of this study is to compare the influence of the water model on the conformational changes of studied proteins in MARTINI simulations. The standard water model (MS),\textsuperscript{37,38} without explicit charges, and the polarizable water model (MP)\textsuperscript{44} that includes two additional particles, bearing partial $\pm q$ charge on the water bead, were used. These charged particles, as mentioned in Chapter 3, allow for mimicking the orientational polarizability of real water molecules and a more realistic electrostatic interaction screening.\textsuperscript{44} The interactions between water beads and charged particles e.g. in the protein, should be the same in the MS and MP simulations, therefore to counter-balance the additional attraction/repulsion of the charged water beads in the MP model, the interactions between beads in MP model have been adjusted to reestablish the interactions of beads in MS model. This has been done by decreasing the van der Waals interactions between charged particles of different water beads (there is an exclusion put on partial charges of the same water bead, and thus they do not interact with each other) and between charged particles of water bead and other bead types in the system. In MS and MP simulations, all setups of saposin-like proteins were prepared using the same protocol, with the exception of used water model. As the strength of interactions in MS and MP models should be on the same level, the observed difference in dynamics of protein in MS and MP simulations must, therefore, arise from the difference in implicit screening. The dielectric constant in the MS model is set up to $\epsilon_{rel} = 15$ and has been decreased in MP model to $\epsilon_{rel} = 2.5$, allowing for stronger interactions between charges via Coulomb potential, which is dependent on the dielectric constant of the medium.

The results from the MS and MP simulations are in qualitative agreement with recent studies\textsuperscript{97,98} showing significant difference between polarizable and standard water models for MARTINI force field in treating interactions between charged amino acid side chains. It was observed that the attraction between unlike charges and repulsion between like charges is underestimated in the standard MARTINI model. This behaviour has been improved with the polarizable water model,\textsuperscript{44} for which the reduced dielectric screening increases the interaction strength between charged amino acid side chains.

5.4.2 Biological Implications

We believe the opening of saposin-like proteins is relevant for the interactions of these proteins with membranes, including the lipid extraction. In presented study we observed full opening of saposin A, C, D and NK-lysin in the MARTINI simulations and partial opening in the atomistic simulations, where opening exposes a hydrophobic core. This would be facilitated in a lipid environment and functionally important. Given the time scales involved in atomistic simulations of protein-lipid interactions and the complex nature of the native membranes/lipids, direct simulation of these proposed steps of the mechanism remains a major challenge, and the MARTINI simulations here and MARTINI simulations of SP-B and lipid aggregates\textsuperscript{21,200,201} suggest that an increase in
simulation effort of 10-100 times compared to the current 2 µs trajectories is likely to be required.

5.5 Conclusions

This study shows that the AMBER99SB-ILDN force field maintains α-helical fold of studied saposin-like proteins, as observed in X-ray and NMR structures, while the GROMOS45a3 force field produces unstable structures that undergo local unfolding. It is known that family of AMBER force field is able to reproduce stable α-structure of proteins, while the GROMOS force field favors turn-coil conformations and extended β-sheet structures. The CD analysis delivers divergent results on the secondary structure content of saposin A, C, D, and NK-lysin so that results from both AMBER and GROMOS simulation could fit the data. If one assumes that the studied proteins should have a high α-helical content, the AMBER force field would be a better choice. On the contrary if one assumes that studied proteins should have a high β-structure elements content, the AMBER force field would show an over-stabilizing effect, and therefore the GROMOS force field would be a more proper choice.

The analysis of the CG simulations revealed significant pattern when comparing MARTINI simulations with standard and polarizable water models. The use of MP model allowed for observing the opening motion of the studied saposins A, C, D, and NK-lysin. Although MARTINI is currently limited in its ability to reproduce protein dynamics, the observed features are likely to be relevant for larger-scale simulations on e.g. relative motions of domains. However, the use of MARTINI with polarizable water model has to be weighed against the higher computational cost compared to simulations with MARTINI with a standard water model.

We believe that the observed tertiary structure changes have important biological implications, as it is expected that the opening of saposins A, C, D, and NK-lysin is crucial for membrane interaction and lipid extraction, and is a prerequisite for function for the whole family of saposin-like proteins. The full opening of the protein is most likely caused by the movement of the stem away from the hairpin, similar to what is observed for the saposin C in the presence of detergent SDS. This flexibility at the opening to a hydrophobic cavity is a feature that is believed to be important for the recognition and extraction of lipids from membranes in many other lipid transport proteins, and thus, could be a feature that is facilitated in lipid environment.
Chapter 6

Conclusions and Future Perspectives
Computational techniques allow us to explore areas in biochemistry that are difficult to investigate with experimental methods. Atomistic molecular dynamics simulations of complex biomolecular systems provide a detailed picture of the studied system, however, it is still challenging to reach time scales longer than a few microseconds for large and complex systems. Consequently, there has been considerable interest in CG models that extend the accessible time scale for molecular dynamics simulations of proteins and their environment, expanding the range of questions that can be addressed by molecular modeling. There are many different CG models varying in the number of atoms which are represented by a single interaction centre, e.g. for proteins some CG models group just a few atoms into a bead, and some group as many as 25 atoms into a single bead.\textsuperscript{32,205} There are some CG models dedicated to study a single type of system, e.g. to study membranes,\textsuperscript{206} or developed to be more generic and versatile, like the MARTINI model.\textsuperscript{37,38,42} Due to a large library of biologically relevant building blocks, such as lipids, sterols, sugars, and amino acids,\textsuperscript{37,38,41,42} the MARTINI model has become very popular, and a good choice to simulate a broad range of different systems. In this dissertation both approaches, modelling using atomistic force fields and CG modeling using the MARTINI force field, have been used to study the dynamics of different proteins on the nanosecond and microsecond time scales, respectively.

In the study of the leucine-binding protein, the MARTINI model showed inability to maintain the structural integrity of the protein, and it is thus necessary to stabilize the tertiary structure of the protein using and elastic network model. An extension to the MARTINI model, the ELNEDIN model,\textsuperscript{40} keeps this structure stable by imposing an elastic network on the leucine-binding protein. However, as ELNEDIN uses a global elastic network, the protein structure is strongly biased to the initial structure, with no chance of describing the protein conformational changes on the microsecond timescale. We therefore proposed the domELNEDIN approach, in which the elastic network is imposed only inside the protein domains, maintaining the structural integrity of the domains while at the same time allowing for their unrestricted movement. The result from ELNEDIN and domELNEDIN simulations indicated that protein dynamics is equivalent in both of these models and comparable to the dynamics of the protein in atomistic simulations on the nanosecond time scale. The situation changes for microsecond time scale simulations, where only domELNEDIN has the potential to describe the protein dynamics, as it allows for unrestricted movements of the protein domains. This study also points out limitations of the MARTINI model. First, since the secondary structure in the MARTINI force field is not flexible, simulations carried out on the leucine-binding protein, and in general on proteins, might suffer from this restriction. A recent study involving the MARTINI model showed that it is possible to introduce flexibility to the the backbone, however this has been done only for specific peptides, and has not yet been transferred to proteins,\textsuperscript{107} which is not a trivial extension. The second limitation concerned the interaction between the protein at study and its ligand leucine. This limitation has been overcome using the 2.2P version of the MARTINI model,\textsuperscript{98} which I tested
during my stay at Professor Siewert-Jan Marrink's group in Groningen, the Netherlands. The new polarizable MARTINI model, version 2.2P, showed significant improvements compared to the previous 2.1 version, in the ability to reproduce the ligand binding to the leucine-binding protein. This model introduces changes only to side chain beads, and therefore no improvement with respect to the flexibility of the backbone of the protein has been achieved. In our studies the closed conformation of leucine-binding protein with and without ligand present is unable to change conformation to the open conformation on the microsecond time scale. Although this could be biologically relevant, it also could show that underestimation of interaction between charged beads, and too high hydrophobicity of a few amino acids in the standard MARTINI model make the protein domains stick to each other. Unfortunately, using the polarizable water model, or the polarizable MARTINI model along with the polarizable water model, did not show any improvements in this matter. We speculate that rescaling the Lennard-Jones interactions could possibly improve this behaviour, however, that would mean that the MARTINI force field should be re-parameterized.

In another study using the MARTINI model, we addressed speculations about the cholesterol influence on the activity of SERCA. In this study a lipid-cholesterol bilayer was allowed to assemble around protein, and possible binding interactions between cholesterol and the transmembrane part of SERCA were spotted. Several annular and non-annular binding sites are observed for the cholesterol molecules, and nothing in our simulations suggests that the probability of finding cholesterol in the lipid annulus is less than finding cholesterol anywhere else in the bilayer, as claimed based on experimental studies. Three structures obtained from CG simulations, with cholesterol in possible non-annular binding modes, will be used for further investigation of the stability of SERCA-cholesterol interaction in atomic detail.

In the study of saposin-like proteins using the MARTINI model, we observed twisting and full opening motions of proteins which, we believe, have important biological implications. It is expected that the opening of saposins A, C, D, and NK-lysin is crucial for membrane interaction and lipid extraction, and this type is speculated to be important for the whole family of saposin-like proteins.

The studies presented in this thesis, show applications of the MARTINI model to different biological problems. Still many aspects of the described projects could be further investigated, using the MARTINI and domELNEDIN models. For instance, for SERCA it is known that anionic lipids affect SERCA activity, and therefore long time scale simulations could be used to elucidate this behaviour. Additionally, self-assembly simulations of bilayers composed of different lipid molecules or mixed bilayers composed of lipids and sterols in different ratios, could be used to mimick the series of fluorescence quenching experiments that had ambiguous results for SERCA. It would also be interesting to perform simulations of the four “corner stone” structures of SERCA from the functional cycle, using the domELNEDIN model, to reveal the full path that the protein
undergoes from one conformation to another. Likewise, another large system that could be interesting to investigate is the saposin C protein inducing vesicle fusion. This has previously been shown possible to study with the MARTINI model in the case of SP-B. This and many more studies could benefit from CG models like MARTINI, ELNEDIN or domELNEDIN.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


Appendix A to Chapter 2

Protein Domain Coarse-Grained Molecular Dynamics: Application to the Leucine Binding Protein

submitted to Journal of Chemical Theory and Computation.
Protein Domain Coarse-Grained Molecular Dynamics: Application to the Leucine Binding Protein

Iwona Siuda,† Lea Thøgersen†,‡,*

†Centre for Membrane Pumps in Cells and Disease, Danish National Research Foundation, Bioinformatics Research Centre, Aarhus University, C. F. Møllers Alle 8, DK-8000 Aarhus C, Denmark

‡CLC bio, Finlandsgade 10–12, Katrinebjerg, DK-8200 Aarhus N, Denmark
ABSTRACT

Periplasmic binding proteins are the initial receptors for the transport of various substrates over the inner membrane of gram-negative bacteria. The binding proteins are composed of two domains, and the substrate is entrapped between these domains. For several of the binding proteins it has been established that a closed-up conformation exists even without substrate present, suggesting a highly flexible apo-structure. For the leucine binding protein (LBP), structures of both open and closed conformations are known, but no closed-up structure without substrate has been reported. Here we present molecular dynamics simulations exploring the conformational flexibility of LBP. Coarse grained models based on the MARTINI force field are used to access the microsecond timescale. We show that a standard MARTINI model cannot maintain the structural stability of the protein whereas the ELNEDIN extension to MARTINI enables simulations showing a stable protein structure and nanosecond dynamics comparable to atomistic simulations, but does not allow the simulation of conformational flexibility. A modification to the MARTINI-ELNEDIN setup, referred to as domELNEDIN, is therefore presented. The domELNEDIN model allows protein domains to move independently and thus allows for the simulation of conformational changes. We show that even though this flexibility is introduced, the structural integrity and nanosecond dynamics are equivalent to that observed in the ELNEDIN and atomistic simulations for LBP. Microsecond domELNEDIN simulations starting from either the open or the closed conformations consistently show that also for LBP, the apo-structure is flexible and can exist in a closed form.
INTRODUCTION

Molecular dynamics (MD) simulations, where all atoms (AA) of a biomolecular system are represented, are well-established and deliver a generous amount of details and insights for the system at study. However, the time scale is limited to hundreds of nanoseconds, and slow dynamics such as conformational changes associated with protein function will typically be on the micro- to millisecond time scale and thus out of reach. Coarse grained (CG) models, where atoms are grouped in “beads” on the amino acid level for proteins,\(^1\)\(^-\)\(^6\) allow for time steps in the 10s of femtoseconds and simulations on the microsecond time scale. The MARTINI CG force field (FF) represents such a model,\(^4\)\(^-\)\(^6\) where the amino acids are described by one to five beads. Within the CG model it is thus possible to distinguish between proteins with similar structure and properties and address questions regarding single point mutations. The microsecond time scale then allows for the modeling of mechanistic events such as membrane insertion\(^7\) and molecular recognition.\(^8\) Protein conformational changes would also in many cases take place within the accessible timescale, however, in these CG models the overall structure of proteins are biased towards the initial structure with harmonic restraints,\(^9\)\(^-\)\(^11\) to avoid a structural collapse during the simulation. For the MARTINI CG FF,\(^4\)\(^-\)\(^6\) the protein secondary structure is stabilized using dihedral restraints on the backbone.\(^5\)\(^,\)\(^6\) This is sufficient for stabilizing the structure of peptides while for larger proteins extra restraints on the structure need to be introduced. One approach for doing this is the ELNEDIN model,\(^11\) where an elastic network (EN) is applied to the whole protein with bonds connecting all CG backbone beads which are within a specific distance in the initial structure. This is of course not optimal, but as can be seen by the popularity of this type of models and by the many new publications applying these methods on interesting
and relevant biological problems, the models are highly applicable for many purposes, despite the current limitations.

In the present study we examine the conformational changes of the periplasmic leucine-binding protein (LBP), using a modification to the MARTINI-ELNEDIN CG method which is applicable to multi-domain proteins. The approach, referred to as domELNEDIN, keeps the residue-level coarse-graining while at the same time allowing the protein to change conformation in an unbiased manner during the MD simulation. ENs are applied only locally in the protein domains, and we show that this is a sufficient degree of stabilization to avoid a collapse of the structure and keep it stable on the nanosecond timescale. As the global nature of the imposed EN is in this way avoided, this model allows for probing the conformational flexibility of proteins.

LBP belongs to the superfamily of binding proteins. More than 100 crystal structures of periplasmic binding proteins (PBP) from different sources and in different conformations have been solved, showing a conserved structural fold, despite high sequence diversity. The protein fold consists of two domains connected by a three-, two- or one-stranded hinge (group I, II and III, respectively). The movement of the domains is a classical example of hinge motion, and the mechanism of ligand entrapment has been referred to as a Venus flytrap model. LBP is a group I PBP, and structures have been established for both open and closed conformations, with the substrate leucine bound in the cleft between the domains in the closed conformation. In gram-negative bacteria, binding proteins found in the periplasm, act as primary receptors in adenosine triphosphate (ATP)-binding cassette (ABC) transport systems, trafficking various substrates across the plasma membrane at the cost of ATP hydrolysis. Despite the great variety in substrates ranging from nutrients such as sugars and amino acids in prokaryotes to polysaccharides, lipids and hormones in eukaryotes, the ABC transport systems seem to share
overall mechanism (for recent reviews see ref. 22-24). For the ABC-importers present in gram-negative bacteria such as *E. coli*, a binding protein entraps the substrate in a closed conformation,\(^{25}\) and docks in this form to the membrane transporter, which is interacting closely with the ATP binding domains on the cytoplasmic site. The ATP-binding and hydrolysis then help fuel the conformational shifts necessary to open the binding protein, release the substrate to the membrane transporter, change the membrane transporter to an occluded state, and finally release the substrate to the cytoplasm.\(^{22}\)

The role of the substrate in the mechanism for the different ABC transporters, as well as the complex interplay between binding protein, conformational changes, and ATP interaction have still not been settled on.\(^{22-24}\) Some of the unsettled issues regarding the substrate role are addressed in the present study of LBP.

The progress of the paper is as follows. First the ability to produce a stable structural scaffold for both the open and closed conformation of LBP is examined on the 100 ns timescale for the standard MARTINI CG, the ELNEDIN and the domELNEDIN models, by comparing the structure and dynamics to atomistic simulations. From this it is evident that the standard MARTINI model is not able to keep the tertiary structure of the protein stable while ELNEDIN and domELNEDIN succeeds in this and furthermore show the same qualities on this short timescale. Following this, microsecond long domELNEDIN simulations are carried out for both the open and closed conformations, to study the conformational flexibility of LBP in water. To illustrate the differences between the ELNEDIN and domELNEDIN models on the long time scale, results from the corresponding ELNEDIN simulations are also reported, and to inspect the sensitivity of the domELNEDIN model towards the assigned protein domain boundaries, four
different domain setups are tested. Finally, the model limitations as well as the biological implications of the results are discussed.

**METHODS**

MD simulations of LBP were carried out starting from a closed (pdb 1USK\textsuperscript{20}) or open (pdb 1USG\textsuperscript{20}) conformation of the protein. For the closed conformation two different setups were applied; either with or without the leucine ligand present in the binding cleft. In all setups counter ions were added (9 Na\textsuperscript{+}) and the protein was solvated with water in a cubic box with dimensions of 100 Å. The GROMACS package version 4.0.7\textsuperscript{26, 27} was used for all simulations and the pressure and temperature were kept constant at 1 bar and 300 K, respectively, using the Berendsen coupling algorithm.\textsuperscript{28} For MARTINI CG simulations it has been shown that the simulated time typically should be multiplied by a factor of 4 to roughly account for the increase in diffusion observed for CG water beads.\textsuperscript{4, 5} Throughout this paper, the simulated time multiplied by 4 will therefore be referred to as “real” time for the CG simulations, and this is the time used in all figures.

**Atomistic Simulations.** The simulations were performed with the AMBER03 FF\textsuperscript{29} for the protein and the SPC water model\textsuperscript{30} for the solvent. Partial charges for the leucine ligand was derived from Antechamber\textsuperscript{31, 32} (see SI Table S1). The temperature and isotropic pressure were kept constant with time constants $\tau_T = 0.1$ ps and $\tau_P = 1$ ps. PME was used for the long-range electrostatics interactions and a cut-off of 10 Å was used for the short-range electrostatics contributions, while the van der Waals interactions were cut off at 14 Å. Bond lengths were constrained using the LINCS algorithm\textsuperscript{33} for the protein. The setups were energy-minimized followed by a relaxation of the solvent and ions, with position restraints (1000 kJ·mol\textsuperscript{−1}·nm\textsuperscript{−2}) applied to all heavy atoms of the protein for 20 ps. Simulations starting from either open or
closed conformations without the leucine ligand present, as well as the closed conformation including the ligand were carried out for 100 ns without any restraints.

**MARTINI CG Simulations.** The simulations were performed with version 2.1 of the MARTINI CG FF. The coarse grained representation of LBP was generated using MARTINI scripts, topologies and parameters. Standard MARTINI CG water beads were used to model the solvent. The leucine ligand was represented by one backbone bead of type P5 and a side chain bead of type C1. The setups were energy-minimized and the solvent and ions were relaxed with position restraints (1000 kJ·mol\(^{-1}\)·nm\(^{-2}\)) applied to all backbone beads of the protein for 1 ns. For the temperature and pressure settings the time constants \(\tau_T = 1\) ps and \(\tau_P = 5\) ps were applied. Non-bonded interactions were cut-off at 1.2 nm and shifted from 0.9 nm for the Lennard-Jones potential and from 0.0 nm for the electrostatic potential. Neighbor lists were updated every 10 steps. Setups with the closed conformation including the ligand and the open conformation without the ligand were simulated for 25 ns (100 ns “real” time) using a 25 fs time step.

**ELNEDIN Simulations.** The applied version of ELNEDIN is based on modifications to version 2.1 of the MARTINI CG FF. In ELNEDIN simulations, an EN is put on the backbone beads of a slightly modified MARTINI CG model of the protein, to maintain the initial tertiary structure. All pairs of backbone beads, for which the distance in the input structure is below some cut-off \(R_C\), are assigned a harmonic network bond with the force constant \(K_S\). Thus, the two input parameters, \(R_C\) and \(K_S\), define the network. For two different setups of LBP (the open form without leucine and closed form with leucine present in the binding cleft) nine different ENs were tested in 25 ns simulations (100 ns “real” time), varying the cut-off distance \(R_C\) (Å) \(\epsilon\) \{8, 9, 10\} and spring force constant \(K_S\) (kJ·mol\(^{-1}\)·nm\(^{-2}\)) \(\epsilon\) \{50, 500, 5000\}. For both setups, the effect of building the model on the protein structure as equilibrated in the atomistic simulations instead of
building it on the crystal structure was also tested for the ENs with parameters \( \{ R_C, K_S \} = \{ 8 \, \text{Å}, 500 \, \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2} \} \) and \( \{ 9 \, \text{Å}, 500 \, \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2} \} \). The equilibrated atomistic structure was found as the average structure over 20 snapshots in the interval 9-11 ns of the atomistic simulations. For the EN providing the best overlap with the atomistic simulations, \( \{ R_C, K_S \} = \{ 8 \, \text{Å}, 500 \, \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2} \} \), 1 μs simulations (4 μs “real” time) were carried out for setups starting from either the closed or open conformations of LBP without ligand present.

All ELNEDIN setups were solvated as in the MARTINI CG simulations. Setups were energy-minimized with position restraints (1000 kJ·mol\(^{-1}\)·nm\(^{-2}\)) and the solvent and ions were relaxed while all protein beads were restrained with the same force constant for 50 ps using 1 fs time step followed by a 1 ns equilibration using 10 fs time step with restraints put only on the backbone beads of the protein. For the temperature and pressure the settings \( \tau_T = 0.5 \, \text{ps} \) and \( \tau_P = 1.2 \, \text{ps} \) were applied. Neighbor lists were updated every five steps. The non-bonded interactions were treated with the same shifts and cut-offs as applied for the MARTINI CG simulations. All simulations were run using a 10 fs time step.

**domELNEDIN Simulations.** There is no unique way to assign each residue in a protein to one or the other structural domain. Four different domain definitions for LBP were therefore tested in this study. Based on visual inspection of both the open and closed conformations it was decided to define domain 1 as residues 1-120 and 250-330 and domain 2 as residues 121-249 and 331-345. This is referred to as the “main” domain assignment. Apart from this, a “main loose” definition, where four amino acids in each of the three domain linkers were released from the ENs was applied, as well as automatic domain boundary assignments from the pDomains server\(^{34}\) and the DomFOLDpdp server,\(^{35}\) as described in the Results section and listed in Table 1.

**Table 1. Protein Domain Boundary Assignments**
Preparation of the setups as well as the production run parameters for the domELNEDIN simulations were the same as for the ELNEDIN model, except the ENs were applied within the protein domains only, as described in Results. For the open form of LBP, and closed form with leucine positioned in the binding cleft, 25 ns simulations (100 ns “real” time) were carried out with the ENs \( \{R_C, K_S\} = \{8 \text{ Å}, 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}\} \) and \( \{9 \text{ Å}, 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}\} \). For the EN providing the best overlap with the atomistic simulations, \( \{R_C, K_S\} = \{8 \text{ Å}, 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}\} \), 1 μs simulations (4 μs “real” time) were carried out for setups starting from either the closed or open conformations of LBP, without the leucine ligand present.

Scripts and example input and output files for converting an ELNEDIN setup to a domELNEDIN setup can be found on http://www.birc.au.dk/~leat/domELNEDIN, and the protocol flow can be seen in SI Figure S1.

**Analysis.** The conformational stability of LBP structures is examined by evaluating the root mean-square deviations (RMSDs) of the protein backbone based on Cα atoms (for the AA simulations) or the backbone beads (for the CG simulations). The RMSD is compared to the structure in the first frame of the simulation, unless stated otherwise. The RMSD and root mean-square fluctuations (RMSF) of individual residues were examined for the last 80 ns of simulation for the AA simulations and last 80 ns of “real” time for the CG simulations, for the protein with the backbone aligned as described above. To compare the large-amplitude fluctuations in the AA and CG (ELNEDIN or domELNEDIN) simulations, the covariance matrix of the positional fluctuations in the AA and CG (ELNEDIN or domELNEDIN) simulations, the covariance matrix of the positional

<table>
<thead>
<tr>
<th>Domain 1</th>
<th>Domain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>main</td>
<td>1-120 + 250-330</td>
</tr>
<tr>
<td>main loose</td>
<td>1-118 + 252-328</td>
</tr>
<tr>
<td>DomFOLDpdp</td>
<td>1-119 + 249-332</td>
</tr>
</tbody>
</table>

#### APPENDIX A
fluctuations was constructed for the coordinates of Cα atoms obtained from the last 80 ns of AA simulations or the backbone beads obtained from the last 80 ns of “real” time of the CG simulations. Trajectories were fitted to the same reference structure, i.e. the X-ray structure of the closed conformation of LBP was used as a reference structure for the simulations starting from the closed conformation of LBP, and the X-ray structure of the open conformation of LBP was used as a reference structure for the simulations starting from the open conformation of LBP. The eigenvectors found when diagonalizing the covariance matrix were then used in the root mean-square inner-product (RMSIP) analysis, that quantifies the overlap between the essential subspaces (described by the 10 first eigenvectors) obtained from the AA and CG simulations: \[ RMSIP = \sqrt{\frac{1}{10} \sum_{i=1}^{10} \sum_{j=1}^{10} (\eta_i^{AA} \cdot \eta_j^{CG})^2} \]

where \( \eta_i^{AA} \) and \( \eta_j^{CG} \) are the \( i \)th and \( j \)th eigenvectors from the AA and various CG (ELNEDIN or domELNEDIN) simulations.
RESULTS

Structure and Dynamics on the Nanosecond Timescale.

Figure 1. Snapshots of the (A) closed and (B) open conformations of LBP at the first frame of the standard MARTINI CG simulation, with corresponding structures after 25 ns of CG simulation (100 ns of “real” time). The protein is colored by residue number going from the N-terminus in red to the C-terminus in blue.

Figure 2. RMSDs for the protein backbone in the AA and standard MARTINI CG simulations starting from the closed and open conformations of LBP.

Standard MARTINI CG. In the MARTINI CG model for proteins each residue is mapped to a backbone bead and zero to four side chain beads, and even though detailed interactions are lost in the coarse graining, the resolution is high enough to represent the particular trades of the different amino acids. The non-bonded parameters are fitted to match the experimentally observed partitioning of amino acids between hydrophobic and hydrophilic environments, and processes driven by this type of interactions should therefore be well described within the model. As directional hydrogen bonds are not possible to represent in the CG model, secondary structure
elements are not self-contained, and the local structure is therefore predefined by restraining backbone angles and dihedrals to values supporting helix or extended structures, based on structural analysis of the initial atomic resolution structure. For purely α-helical structures or membrane proteins completely surrounded by lipid membrane, MARTINI CG simulations have been reported where no further restraints on the structure have been imposed. However, β-sheet interactions are non-local in nature and therefore poorly described by local restraints, and as water does not stabilize structure in the same manner as a lipid bilayer, the standard MARTINI CG model does generally not succeed in maintaining the tertiary structure of globular proteins. The MARTINI CG simulations of LBP in water are no exception. For both the simulation starting from the open and the one starting from the closed conformation of the protein, the tertiary structure is observed to collapse into a packed globular structure in an unspecified manner (see Figure 1-Figure 2), which does not resemble the known structure of the protein in a closed conformation. The collapse of tertiary structure resulting from MARTINI CG simulations has been the main motivation for the previously proposed ELNEDIN extension to the MARTINI FF as well as the modification, domELNEDIN, presented in this work. The inability of the standard MARTINI CG model to keep the structure stable for even 100 ns makes this model irrelevant for further studies of LBP, and instead the focus will be on the above mentioned extensions to the MARTINI CG model, where the tertiary structure of the protein is stabilized by the application of an EN to the backbone beads.

The ELNEDIN Extension. The ELNEDIN extension to MARTINI CG was intended to stabilize the overall protein structure, while at the same time allowing for structural fluctuations on the nanosecond timescale, comparable to that observed for AA simulations. This was achieved through the application of an EN to backbone beads of a MARTINI CG model. Minor
modifications to the standard MARTINI CG model were introduced with the ELNEDIN extension, as the backbone beads now were placed in the Cα positions instead of at the center-of-mass of the backbone atoms. Two parameters were used to define the network and tune the dynamics, namely the cut-off distance, $R_c$, for applying an EN bond between two beads, and the force constant, $K_s$, of the harmonic EN bonds forming the network.

For the purpose of establishing the optimal set of parameters to describe the dynamics of LBP using the ELNEDIN model, simulations were carried out with all possible combinations of three different force constants and three different cut-offs, as described in the Methods section. The choice of the scaffold parameters to use for further simulations is made based on a comparison with the protein dynamics observed in an atomistic simulation. Just as in the original ELNEDIN study, the RMSD as a function of time and the RMSD per residue are used to quantify the global and local structural deformations, while the RMSF per residue and the essential subspace overlap (quantified by the RMSIP) are used to describe the local and large-amplitude fluctuations, respectively. As expected, the RMSD and RMSF values decrease in a systematic manner with the increase of the cut-off and force constant values (SI Figure S2-Figure S7). The RMSIP cannot be expected to show such a systematic behavior, but in this case the overlap consistently increases when increasing the cut-off and force constant (SI Table S2 and S3). As LBP is a protein with 345 residues, an estimated lower-end RMSIP is found as $\sqrt{10/(3 \times 345 - 6)} = 0.099$. The RMSIP values are in all cases found to be between 0.50 and 0.75, and the overlap is thus highly satisfactory for all the tested parameter sets. From Figure S2-S7 it is seen that the parameters giving the best overlap with the atomistic data is $K_s = 500 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$, suggested as default value for the ELNEDIN model, and a cut-off of $R_c = 8 \text{ Å}$. The cut-off suggested as
default for the ELNEDIN model is $R_C = 9 \text{ Å}$, and the effect of changing the cut-off by 1 Å is also seen in Figure 3 for both the open and closed conformation.

Figure 3. RMSD as a function of simulation time and RMSF and RMSD per residue for the last 80 ns of simulation obtained from ELNEDIN simulations of the closed conformation of LBP with ligand (A) and the open conformation without ligand (B). The force constant parameter, $K_S$, of the EN is in all cases 500 $\text{kJ mol}^{-1} \text{nm}^{-2}$ and the cut-off, $R_C$, is either 8 Å or 9 Å as indicated on the figure. Plots include results for simulations starting from either the crystal structure (X-ray, red line) or a structure equilibrated in an atomistic simulation (AAeq, orange line).

As the bonds forming the EN in ELNEDIN are put on the initial structure, and the original distances are used as the equilibrium distances for the EN bonds, the simulation will by construction be biased towards the original structure, and no major structural changes can take place. When applying an EN it is thus important that the initial structure represents a stable conformation. However, as it is not possible to optimize the structure within the unbiased MARTINI CG FF as discussed above, the choice is either to use directly the crystal structure of
the protein or to carry out a short equilibration in atomistic resolution before imposing the EN bonds. The changes in the structure resulting from an AA MD equilibration will not only change the equilibrium distances of the EN bonds, but it will also change which bonds are formed as some bead distances will get below cut-off and vice-versa. From Figure 2 it is seen that the backbone RMSDs of both the open and the closed conformations of LBP have reached a plateau already after 9 ns of AA simulation. To examine the sensitivity of the sampled dynamics to the choice of initial structure, simulations were therefore carried out where the ELNEDIN model was either based on the crystal structure (X-ray) or an equilibrated structure obtained from the atomistic simulation (AAeq), as described in the Methods section. The results are shown in Figure 3 and Table S4 in the SI. If the structure had been relaxed in a FF identical to the modified MARTINI used in the ELNEDIN model, it would be expected that the relaxed case would show to be more stable in the ELNEDIN simulations than the setup starting from the crystal structure. However, as in this case the structure has been relaxed within an AA FF prior to the CG mapping, the outcome was not a given. From the RMSD graphs in Figure 3, it is seen that the overall change of the structure observed during the ELNEDIN simulations is on the same level no matter the choice of starting structure, and so are the RMSIPs (SI Table S2-S4). For the RMSF and the RMSD per residue, a significant difference is seen depending on the initial structure, most prominent for the open conformation. The overall EN is much denser for the 9 Å cut-off (~1700 EN bonds) than for the 8 Å cut-off (~1100 EN bonds), and the effect of the chosen initial structure is therefore more pronounced for $R_C = 8$ Å. As a general trend the simulations starting from the crystal structures show either equivalent or dampened dynamics compared to the ones starting from the structures equilibrated in atomistic simulation. The backbone RMSD between the crystal structure and the average structure is 1.5 Å for the closed
conformation and 2.7 Å for the open conformation. These differences in structure result in ENs where 7-10% of the EN bonds differ between the models based on the different initial structures (Figure 4). Furthermore, the ENs based on the crystal structures have from 12 to 44 EN bonds more than their counterparts starting from the equilibrated structure. The sensitivity towards the choice of initial structure that here presents for the ELNEDIN model is troubling but not surprising as this is a fundamental issue for elastic network model based methods. It is clear that for LBP a preceding equilibration of the structure in the AMBER atomic resolution FF does not result in an ELNEDIN model which compares better to the nanosecond dynamics in the atomistic simulation than by simply starting from the crystal structure, in fact rather on the contrary.

Figure 4. Comparison of ENs generated for the ELNEDIN model based on the crystal structures (X-ray) and equilibrated structures obtained from the atomistic simulation (AAeq). The EN scaffold for the closed form with ligand (A) and the open form without ligand (B) was parameterized with the cut-off $R_c = 8$ Å. Bonds are colored as follow: gray – bonds present in both “X-ray” and “AAeq” setups, red – specific bonds for the “X-ray” setup, blue – specific bonds for the “AAeq” setup.

As previously documented, the ELNEDIN extension to the standard MARTINI CG model succeeds in reproducing a structural scaffold similar to what is seen in atomistic simulations of
For the case of LBP this is achieved building the EN based on the protein crystal structure and using the parameters $R_C = 8 \, \text{Å}$ and $K_S = 500 \, \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$. It should be recognized as a quality of the model that, despite quite different overall levels of dynamics, the ELNEDIN setup giving the best fit to the atomistic simulations is the same for both the closed and open conformations. A major drawback, though, is that the structural scaffold is limited to a single protein conformation, not allowing studies that involve rearrangements of the protein domains. To circumvent this bias towards the initial conformation, we have examined the effect on the protein structure and dynamics when removing the EN bonds between protein domains.

The domELNEDIN Extension. There is no unique definition of how to divide a protein into domains, and various criteria have been used such as structure, function, folding units, sequence or evolution. In the growing number of known protein structures, domains containing highly conserved combinations of secondary structure elements repeated in different proteins or within the same protein are observed. The structural difference between different protein conformations is seen mainly to arise from hinge, shear, or rotational motions between these structural domains. In the case of LBP, where the structure of both an open and closed conformation is known, it is clearly a two-domain protein. The RMSD between the crystal structures of the open and closed conformation of LBP used for this work is 7.0 Å, while the RMSD between the same domains of these two conformations are 0.7 Å for domain 1 and 0.6 Å for domain 2, when using the main domain assignment (Table 1). The structural rearrangements between the conformations are thus mainly caused by movements in the hinge region of the protein. For other globular proteins, where we have compared crystal structures for different conformations, it is also clear that structural domains can be assigned which show a stable core no matter the conformation. For LBP and many other multi-domain proteins, the individual
protein domains could therefore be restrained with EN bonds as in an ELNEDIN model, while still allowing for a full description of the structural ensemble covering the whole functional cycle. This is the concept of the domELNEDIN model; an EN is applied within each domain with the same type of parameter setup as for ELNEDIN, the difference being that no EN bonds are applied connecting the protein domains. For LBP, the differences between ELNEDIN and domELNEDIN are seen in Figure 5 for both the open and the closed conformations. In order for the domELNEDIN model to be successful, the non-bonded interactions between protein domains should be well described by the MARTINI CG FF. MARTINI has previously successfully been used to study macro-assemblies of proteins\textsuperscript{11} and the aggregation of proteins,\textsuperscript{11, 14, 39, 44, 45} and recently a thorough study of side chain dimerization showed that the MARTINI FF produced dimerization free energies which in aqueous solution correlated reasonably with those obtained with the atomic resolution FFs OPLS\textsuperscript{46, 47} and GROMOS.\textsuperscript{48, 49} It would therefore seem reasonable to expect that the correct packing and interaction of protein domains can be described within the FF as well, as domain packing holds many similarities to protein-protein interaction.

![Figure 5](image_url)

Figure 5. Comparing the ELNEDIN and domELNEDIN models for ENs generated with $R_C = 8\ \text{Å}$, and using the main domain assignment. (A) Closed conformation. (B) Open conformation. Bonds are colored as follow: gray – bonds present in both the ELNEDIN and domELNEDIN models, red – bonds not included in the domELNEDIN model.
Just as for the ELNEDIN model, simulations were carried out to establish the most suitable set of EN parameters for a domELNEDIN simulation of LBP. If the annihilation of the inter-domain EN bonds does not completely destabilize the structure, it is to be expected that the dynamics on the 100 ns time scale is similar for the ELNEDIN and domELNEDIN models, and therefore only the parameters $K_S = 500$ kJ mol$^{-1}$ nm$^{-2}$ and $R_C = 8$ Å and 9 Å were inspected using the X-ray structures as initial structures for the modeling. As the protein domains are allowed to move independently, it is not expected that the dynamics will be dampened in a systematic manner going towards a higher cut-off and force constant. Still, the overall best parameter fit is also for domELNEDIN achieved when applying the cut-off $R_C = 8$ Å (SI Figure S8 and Table S5). In Figure 6, the RMSD and RMSF data are compared for the ELNEDIN and domELNEDIN simulations, and in Table S5 in the SI, the RMSIP values are compared. It is clear that even though the overall RMSD shows a more flexible protein structure for the domELNEDIN simulations, the RMSD and RMSF per residue are very similar between the two models, and as well are the RMSIP values. The ELNEDIN and domELNEDIN extensions to the MARTINI CG FF thus show the same level of structural stability on the 100 ns time scale while using the same EN parameter set. This stability is achieved for the domELNEDIN model despite the removal of the EN bonds stabilizing the domain interfaces, and thus conformational changes are allowed in an unbiased manner within this model. The difference between the ELNEDIN and domELNEDIN models should therefore be noticeable and important on the microsecond time scale.
Figure 6. Backbone RMSD as well as RMSF and RMSD per residue in the ELNEDIN and domELNEDIN simulations of closed form with ligand (A) and open form without ligand (B) of LBP. The EN scaffolds were parameterized with $R_C = 8$ Å and $K_S = 500$ kJ·mol$^{-1}$·nm$^{-2}$.

**Long Time Scale Events.**

For a number of the binding proteins it has been established that the apo-form is flexible and occupies a wide range of conformations ranging from a full opened to a closed conformation,$^{50-56}$ and the open-to-closed transition is expected to take place on the nanosecond to microsecond timescale.$^{50}$ A closed conformation without substrate has not yet been observed for LBP, and to study the conformational flexibility of LBP when substrate is not present, microsecond long domELNEDIN simulations have been carried out, starting from both the open and closed conformations. For both cases, the domELNEDIN simulations are compared to the corresponding ELNEDIN simulations. Furthermore, the sensitivity towards the domain boundary definitions and towards the choice of initial structure to use for the domELNEDIN setups are tested.
Starting From the Open Conformation. Figure 7, the development of the protein backbone RMSDs over the full simulations are depicted compared to the crystal structures of both the open and the closed conformations. The ELNEDIN model applied to the open conformation is expected to result in simulations producing a structural ensemble around the crystal structure of the open conformation. However, on the contrary a change towards the closed conformation is observed on the microsecond time scale, going from an RMSD of 7.0 Å to 5-6 Å with respect to the crystal structure of the closed conformation. As clear from Figure 5B, there are only a few EN bonds connecting the domains in the open conformation (red lines), and they are all positioned in the hinge region. This is why an overall structural change is allowed without conflicting too much with the globally applied EN. However, the conformational change has a limit, as further change towards the closed conformation will be energetically highly unfavorable due to the necessary change in the EN bonds bridging the domains.

Figure 7. Backbone RMSDs in ELNEDIN and domELNEDIN simulations starting from the open conformation of LBP. (A) Compared to the crystal structure of the open conformation. (B) Compared to the crystal structure of the closed conformation. Left-most graphs show results
from the ELNEDIN simulation, the domELNEDIN simulation using the main domain assignment and the “topology swap” setup, where the topology for the closed conformation is applied to the domELNEDIN simulation starting from the open conformation. The right-most graphs show results from domELNEDIN simulations using three alternative domain assignments in the setup. The protein domain boundaries corresponding to main, main loose, pDomains and DomFOLDpdp can be seen in Table 1.

In the domELNEDIN simulations the open conformation is free to close up, and for all setups the structure clearly approaches the closed conformation, and to a significant higher degree than seen for the ELNEDIN simulation, resulting in RMSDs as low as 3 Å with respect to the crystal structure of the closed conformation, starting from an RMSD of 7 Å.

Even though the protein domains are structurally very similar between the conformations, there are a few differences in their modeling parameters since equilibrium angle values for the backbone as well as the EN bonds are assigned based on the exact atom positions in the initial structures. Out of the ~1080 bonds forming the EN for the domELNEDIN simulations ~50 are unique for the closed or open structures (around 10 of these involve residues in the domain linkers) while the rest are shared between the setups for the two conformations. A setup was made named “topology swap”, where the topology input for the closed conformation was applied to the open conformation, enforcing the EN bonds and equilibrium backbone angles of the closed conformation. For this setup, a simulation starting from the open conformation is seen to reach a closed structure which has an RMSD as low as 1.6 Å compared to the crystal structure of the closed form. The application of the structural parameters from the known closed structure thus allowed the proper ”induced fit” of the domain structures into the closed conformation of the protein.
The division of a protein into structural domains will often be done in the most sensible way by a trained eye, and the well-established fold databases CATH and SCOP rely on human expertise. However, with the pace at which new protein structures are submitted to the protein data bank (PDB), the human expert assignments lag behind, and the development of methods for high-quality automatic assignment of domain boundaries is an active field of research. To inspect how the outcome of the domELNEDIN simulations of LBP depends on the domain boundary assignment, four different assignments have been tested (Table 1). The main domain assignment was based on our judgment from visual inspection of both the closed and open conformations, and it is identical to the assignment by CATH. The “main loose” assignment has the same boundaries as main, except four amino acids in each of the three domain linkers are released from the ENs. This setup was chosen to test if a more flexible linker region would alter the outcome. Two alternative protein domain boundaries were also acquired from two public available servers for automatic protein domain assignment. One is the DomFOLDpdp server, which based on the amino acid sequence predicts the protein fold using the nFOLD3 method, and then analyze the best fold model with the Protein Domain Parser (PDP) program. The other is the pDomains server, which for a protein PDB Id can provide an overview of domain assignment results from seven different methods, as well as provide a consensus assignment based on these results. In the consensus assignment, the different methods contribute with a weight which is based on benchmarked knowledge of each methods performance for that particular type and size of protein. For the closed conformation two consensus assignments were presented, the first consensus assigned it as a one-domain protein while the second consensus was a two-domain assignment. The two-domain consensus assignments for the open and closed structures used in this study were identical, except residue 333 was assigned to
domain 1 in the closed conformation and domain 2 in the open conformation. As seen in Table 1, we used the consensus assignment as given to the open conformation.

The four simulations using different domain assignments show the same overall result; in all cases the LBP starts from an open conformation and ends up in a closed conformation after 4 $\mu$s of simulation. However, as could be expected, the highest degree of flexibility is observed for the “main loose” setup. For this setup a number of structural changes between open and closed-like conformations are seen on the 4 $\mu$s time scale, while the three other setups only show a single change towards the closed conformation in the same time frame. All four simulations differ in the time it takes to make the change to the closed conformation, but to conclude if the fast conformational change observed with the pDomains assignment contrary to the much slower conformational change observed with the main assignment are inherent to the domain assignments, and not a result of the stochastic nature of the event, would require multiple repeat simulations of the same setup. The presented simulations show that one should, as always, be very careful when concluding anything relating to the simulation time based on a single simulation. Furthermore, using the domELNEDIN model, the dynamics of the domain movements will very likely be dependent on how tight the linkers between the domains are bound to the domains themselves.

Starting From the Closed Conformation. In all simulations of the closed conformation LBP stays more or less closed as seen in Figure 8, even though the ligand was not present. In the ELNEDIN simulation it is under no circumstances expected to observe any opening of the protein, as several EN bonds are applied between the domains, keeping them tightly together (Figure 5A). Also for the domELNEDIN simulations the closed conformation is very stable and no opening is observed. Thus, even though the protein is free to change its conformation in a
long time scale simulation, it stayed in the stable closed conformation. A 100 ns AA simulation starting from the closed conformation without ligand present was also carried out (data not shown), and it showed consistently that the ligand-free structure was highly stable (RMSD of $1.5 \pm 0.1 \ \text{Å}$ over the last 50 ns), and even more so than when the ligand was present (RMSD of $1.9 \pm 0.2 \ \text{Å}$ over the last 50 ns – Figure 2).

Figure 8. Backbone RMSDs in ELNEDIN and domELNEDIN simulations starting from the closed conformation of LBP. (A) Compared to the crystal structure of the closed conformation. (B) Compared to the crystal structure of the open conformation. Left-most graphs show results from the ELNEDIN simulation, the domELNEDIN simulation using the main domain assignment and the “topology swap” setup, where the topology for the open conformation is applied to the domELNEDIN simulation starting from the closed conformation. The right-most graphs show results from domELNEDIN simulations using three alternative domain assignments in the setup. The protein domain boundaries corresponding to main, main loose, pDomains and DomFOLDpdp can be seen in Table 1.
Similar to the “topology swap” setup for the open conformation, a setup was made where the topology input for the open conformation was applied to the closed conformation, enforcing the EN bonds and equilibrium backbone angles of the open conformation onto the closed structure. In this simulation, the protein clearly remained closed, while remodeling the structure locally, to better fit with the EN bonds and equilibrium angles specific for the open conformation, thus resulting in an elevated RMSD to the closed structure while keeping an even higher RMSD to the open structure. This observation corroborates that the application of the topology from a different conformation does not force a conformational change in itself, but merely introduce the induced fit of the internal domain structure corresponding to the opposite conformation.

The different domain assignments do also in this case not alter the outcome of the simulations. Even the release of the linkers from the ENs of the protein domains in the “main loose” setup does not affect the overall structural changes within the 4 µs time frame.

**DISCUSSION**

As both standard MARTINI CG and ELNEDIN simulations do not allow the study of conformational flexibility, the discussion only concerns the AA and domELNEDIN simulations.

**Flexible Apo-structure Moves from Open to Closed Conformation.**

For several of the binding proteins, it has been established that the apo-structure is very flexible, with a structural diversity going all the way from a completely open to a closed-up conformation\(^{50-56}\) and with an open-to-close exchange on the microsecond timescale.\(^{50}\) This is also what we observe for LBP. Both AA and domELNEDIN simulations, starting from the open conformation, show a highly dynamical structure (Figure 6B), which fluctuates around a fully open conformation on the 100 ns timescale. Then, in the domELNEDIN simulations extending to the microsecond timescale, conformational changes towards a closed conformation is clearly observed (Figure 7). For those of the binding proteins, where structural information on a closed
conformation without ligand has been established, cases where the ligand bound and unbound structures are almost identical are observed\textsuperscript{52,56} as well as a case where the closed conformation without substrate seems to close up in a distinguishable different conformation than when the substrate is present.\textsuperscript{50} In the domELNEDIN simulations, the closed-up structure obtained from the simulations starting from the open conformation has a backbone RMSD of around 3 Å (Figure 7B) compared to the crystal structure of the closed conformation with ligand bound. The degree of closing of the open apo-structure could be dependent on the exact domain definitions applied in the setup, but it seems not to be (Figure 7). That the RMSD does not go below 3 Å can either be because LBP actually closes up in a different manner when leucine is not present, or be due to limitations in the coarse domELNEDIN model.

A limitation of the domELNEDIN model is that the ENs set up inside the protein domains are based on the original structure, and any induced fit going from an open to a closed structure will thus not be supported. As the RMSDs between the domains in the closed and open structures are 0.7 Å for domain 1 and 0.6 Å for domain 2, this effect was not expected to be significant. However, if the simulation starting from the open conformation is applied the closed structure topology (topology swap), and thus enforced the equilibrium distances between backbone beads inside the domains corresponding to the closed structure, LBP was observed to undergo a conformational change all the way from the open conformation to a structure with RMSD of only 1.6 Å with respect to the crystal structure of the closed conformation (Figure 7B). This shows that induced fit of the domains indeed play a role in the complete close-up of the protein, but from these simulations it cannot be determined if this fit is induced by interaction with the substrate or if it could also take place in the apo-situation. To address this, atomistic simulations
are needed, e.g. starting from a reverse coarse graining of the closed up apo-structure observed in the domELNEDIN simulations, and both with and without a leucine ligand included in the setup.

**Highly Stable Closed Conformation.**

The simulations starting from the closed conformation with the leucine ligand removed from the setup show a very stable structure, also on the microsecond timescale (Figure 8A). Even when the topology for the open conformation is applied in the simulation starting from the closed conformation, the structure keeps a stable closed conformation, although the internal domain structure and backbone are changed to match those in the open conformation.

When interpreting the results, it should be kept in mind that the simulations starting from the closed apo-structure are artificial in the sense that the crystal structure used for the setup contained the substrate. It could be that an energetic barrier would keep the structure from closing this tightly in vivo without leucine present, and the leucine ligand would induce a fit in the structure, as discussed in the previous sub-section.

Also, it cannot be ruled out that the observation of the microsecond stability of the closed conformation is due to an over-stabilization of the protein-protein interaction in the MARTINI CG FF. However, AA simulations of the closed ligand-free conformation also show a very stable structure, albeit at the 100 ns timescale, and, as described in the Results section, MARTINI has previously successfully been used to study protein-protein interactions,11, 14, 39, 44, 45, 49 and it therefore seems reasonable to expect that the interaction of protein domains can be described within the FF as well. Naturally, the coarse model is a compromise, and in a very recent update to the MARTINI protein model (version 2.267), the description of side-chains have been further improved by adding particles with opposite charges on polar side chains as well as moving the charge on charged side-chains away from the van der Waals center of the charged bead, to allow
for a higher resolution in the modeling of electrostatics interactions. The water model has also been extended with particles with opposite and movable charges, to allow for polarization of the water beads. Using this new augmented MARTINI CG protein and water model together with domELNEDIN on the closed conformation with the substrate removed, microsecond long simulations consistently show a stable structure also in this improved CG FF (data not shown).

All microsecond domELNEDIN simulations in this study agree that LBP can close up independent of leucine presence, as observed for several other binding proteins. Whether this closed-up conformation is identical to the one observed for the substrate-bound state, or if it deviates with an RMSD of 3-4 Å, as seen in the simulations starting from the open conformation, the existence of a closed conformation without substrate has implications for the ABC-transporter mechanism, that we will consider in the following.

Figure 9. Schematic of leucine transport through the membrane. The mechanistic cycle involved when LBP closes up without substrate bound, as suggested by the present study, is also included, and gray arrows indicate where it differs from the substrate-transport cycle. (I) LBP in the open,
substrate-free form (dark gray) and unbound substrate (black) in the periplasm. (II) The substrate binds to LBP. (III) LBP closes up. (IV) LBP docks to the transmembrane permease (TMP) (light gray), which is connected with the ATP-binding domains (ATP_BD) on the cytoplasmic side. (V) Binding triggers the opening of LBP and release of the substrate to the permease, if substrate is present, as well as release of the binding protein back to the periplasm. (VI-VII) The substrate is transported through the permease and released into the cytoplasm. (VIII) The permease is back in a state ready to interact with a closed up binding protein.

The binding protein interacts with a transmembrane permease, and in this way helps ensure that the substrate transport is unidirectional.69 In the few crystal structures of full ABC-transporter complexes, where the interaction between a transmembrane permease and its binding protein can be seen,70-73 the two subunits of the permease interact with each their domain of the binding protein. It therefore seems plausible that the binding protein in its closed form docks to the permease while it is in an inward facing or occluded conformation (Figure 9 IV). An ATP-driven conformational change to an outward facing form of the permease would then break open the binding protein, for the substrate to be released into the permease (Figure 9 V). Yet another conformational change would convert the permease to an occluded conformation, releasing the binding protein in its open form into the periplasm again (Figure 9 VI).73 The permease interaction would by this mechanism help the binding protein overcome the energy barrier going from the closed to the open conformation, ensuring that the ligand is not released erroneously when first captured by the binding protein.

At a first glance, this mechanism seems unlikely to be combined with the presence of a closed conformation without substrate, as this would result in seemingly futile ATPase activity.
Nonetheless, there are also experimental data and observations that support the idea of substrate independent interaction between the binding protein and the permease. The kinetics of a binding-protein dependent transport system has been modeled, and it is clear that a model where only the substrate bound binding protein is recognized by the permease does not explain the experimental data, whereas a model where the substrate-bound and empty binding proteins compete for the permease interaction does.\textsuperscript{74, 75} It has been shown that the histidine binding protein has equal affinity for the transmembrane permease whether loaded with substrate or not,\textsuperscript{76} and that the binding protein without substrate competes efficiently with the loaded binding protein, and thus inhibits the substrate transport.\textsuperscript{77} The ATPase activity is also stimulated by the substrate-free binding protein, albeit at a lower level than when substrate is present.\textsuperscript{76} Similar measurements have been carried out for the vitamin B12 importer. In this case the ATPase activity was equally stimulated by the binding protein with or without substrate, which suggests that the transporter is unable to distinguish between empty or B12-loaded binding protein.\textsuperscript{78} These observations point to an overall inefficient substrate transport, as ATP hydrolysis does not seem tightly coupled to substrate transport. However, it has been recognized that other ATP-burning machines operate at efficiencies much lower than 100 \%,\textsuperscript{79, 80} and it could be that it is rather an unlikely scenario to have biological machines running at maximal efficiency.

**Substrate Dependent Conformational Changes.**

The interactions between LBP and the leucine ligand are of hydrophobic, hydrogen-bonding and electrostatic character. A proper modeling of the zwitter-ionic nature of the “backbone” part of the leucine ligand is therefore necessary to model the substrate effect on the conformational changes. Within the MARTINI 2.1 CG FF the free leucine ligand is based on ref. 6 modeled as two beads with the side chain part mapped to a bead of type C1 (apolar with lowest polarity
level; 1 out of 5) and the backbone part to a bead of type P5 (polar of the highest degree), and it is not possible to add multiple charge points to this “backbone” description. This causes the interaction between the closed protein and the leucine ligand to break after a few hundred nanoseconds in both ELNEDIN and domELNEDIN simulations, without any noticeable effect on the protein structure (data not shown). In the newest version of MARTINI (v. 2.2), the possibility for modeling beads with multiple charge points have been introduced, and work is in progress to describe the LBP-substrate interaction within this updated FF, to also study substrate dependent conformational changes.

CONCLUSION

Using MD simulations we have studied the conformational flexibility of LBP when the substrate leucine is not present. CG models are used to access the dynamics on the microsecond timescale, and in this study three CG models based on the MARTINI CG FF have been evaluated for the purpose. The standard MARTINI CG FF is shown not to maintain a stable structure at the 100 ns timescale. The ELNEDIN model, where harmonic bonds are added between close-by backbone beads forming a global EN, is shown to provide simulations with a stable structure as well as dynamics comparable to atomistic simulations on the 100 ns timescale. However, due to the EN, this model does not allow the description of conformational flexibility. The domELNEDIN extension, presented in this work, is a modification to ELNEDIN where all EN bonds connecting protein domains are left out while keeping the bonds inside the protein domains. The domELNEDIN simulations of LBP show the same qualities as ELNEDIN on the 100 ns timescale, reproducing both the structural stability and the nanosecond dynamics comparable to atomistic simulations, while at the same time allowing for unbiased movements of the protein domains.
In microsecond long simulation, four different ways of assigning the 345 residues of LBP to the two protein domains were then tested, and the outcome of the domELNEDIN simulations was seen not to depend significantly on the exact domain assignment. As the formation of harmonic bonds in the ENs is based on interatomic distances in the initial structure used for the setup, the domELNEDIN model (topology) established from respectively the open and closed conformations of LBP differ. Applying the closed conformation domELNEDIN model in a simulation starting from the open conformation and vice versa can therefore enable the modeling of a more complete conformational change. It is in this way observed that for LBP, an induced fit is needed for the open conformation to close up completely.

It could also be imagined that the introduction of a non-uniform stiffness of the ENs would allow for the simulation of induced fit between protein domains, e.g. by making the network weaker in areas based either on distance to the protein surface or based on knowledge from a short AA simulation. Structures from multiple protein conformations could also be used to derive a common domELNEDIN model for a particular protein, where the differences in the domELNEDIN models established from the individual structures are removed or modified, to allow for a model which is not biased by one structure in particular. It has to be tested, though, if such setups would still produce dynamics comparable to atomistic simulations on the nanosecond timescale.

For several of the binding proteins, being part of the ABC transport system in gram-negative bacteria, it has been established that a closed-up but empty conformation exists, and all microsecond domELNEDIN simulations presented in this study agree that this is also the case for LBP. Due to the coarseness of the model, it cannot be determined if the closed empty
conformation is identical to the substrate-bound conformation, or if the substrate induces a fit in the structure.

As the conformational changes guarding the transport of the substrate from its binding protein and through the transmembrane permease are ATP-driven, the existence of almost identical empty and substrate-bound binding proteins should result in seemingly futile ATP hydrolysis. Our results thus support previous experimental observations, which suggest that the ABC-transport system is not exclusively triggered by a substrate-bound binding protein.\textsuperscript{76-78}

We plan to use the domELNEDIN approach together with the latest MARTINI CG FF improvements, to also model how the conformational flexibility of LBP is affected by the presence of substrate. As structures of full ABC-transporter complexes are known, it would also be interesting to model how the stability of the closed conformation of a binding protein with and without the substrate bound is affected by the interaction with the transmembrane permease.

ASSOCIATED CONTENT

Supporting Information. Supplemental Information includes seven figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Correspondence should be addressed to: lea@birc.au.dk

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.
ACKNOWLEDGMENT

This work was supported by funds from the Danish National Research Foundation for the Pumpkin Centre, the Carlsberg Foundation (LT), and from the Aarhus Graduate School of Science and Technology (IS).

REFERENCES


60. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.;


64. Jones, D. T.; Bryson, K.; Coleman, A.; McGuffin, L. J.; Sadowski, M. I.; Sodhi, J. S.;


DOI: 10.1021/ct300646g.


Supporting Information

Protein Domain Coarse-Grained Molecular Dynamics: Application to the Leucine Binding Protein

Iwona Siuda,† Lea Thøgersen†,‡,*

†Centre for Membrane Pumps in Cells and Disease, Danish National Research Foundation, Bioinformatics Research Centre, Aarhus University, C. F. Møllers Alle 8, DK-8000 Aarhus C, Denmark

‡CLC bio, Finlandsgade 10–12, Katrinebjerg, DK-8200 Aarhus N, Denmark
<table>
<thead>
<tr>
<th>Atom type</th>
<th>Atom name</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>amber99_39</td>
<td>N</td>
<td>-0.066</td>
</tr>
<tr>
<td>amber99_17</td>
<td>H1</td>
<td>0.238</td>
</tr>
<tr>
<td>amber99_17</td>
<td>H2</td>
<td>0.238</td>
</tr>
<tr>
<td>amber99_17</td>
<td>H3</td>
<td>0.238</td>
</tr>
<tr>
<td>amber99_11</td>
<td>CA</td>
<td>-0.218</td>
</tr>
<tr>
<td>amber99_28</td>
<td>HA</td>
<td>0.128</td>
</tr>
<tr>
<td>amber99_11</td>
<td>CB</td>
<td>-0.173</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HB1</td>
<td>0.125</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HB2</td>
<td>0.125</td>
</tr>
<tr>
<td>amber99_11</td>
<td>CG</td>
<td>-0.121</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HG</td>
<td>0.093</td>
</tr>
<tr>
<td>amber99_11</td>
<td>CD1</td>
<td>-0.213</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD11</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD12</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD13</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_11</td>
<td>CD2</td>
<td>-0.213</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD21</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD22</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_18</td>
<td>HD23</td>
<td>0.075</td>
</tr>
<tr>
<td>amber99_2</td>
<td>C</td>
<td>0.354</td>
</tr>
<tr>
<td>amber99_45</td>
<td>OC1</td>
<td>-0.492</td>
</tr>
<tr>
<td>amber99_45</td>
<td>OC2</td>
<td>-0.492</td>
</tr>
</tbody>
</table>

Table S1. Atomic charges for the zwitter-ionic leucine ligand generated using Antechamber.

<table>
<thead>
<tr>
<th>$K_S$ / kJ/(mol·nm²)</th>
<th>$R_c$/Å</th>
<th>50</th>
<th>500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.50</td>
<td>0.63</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.53</td>
<td>0.65</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.55</td>
<td>0.67</td>
<td>0.68</td>
<td></td>
</tr>
</tbody>
</table>

Table S2. RMSIP values for simulations of the closed conformation of LBP with ligand, quantifying the essential subspace overlap between the AA simulation and the ELNEDIN simulations varying the cut-off distance, $R_c$, and spring force constant, $K_S$. Obtained from the last 80 ns of “real” time of simulation.
Table S3. RMSIP values for simulations of the open conformation of LBP, quantifying the essential subspace overlap between the AA simulation and the ELNEDIN simulations varying the cut-off distance, $R_c$, and spring force constant, $K_S$. Obtained from the last 80 ns of “real” time of simulation.

<table>
<thead>
<tr>
<th>$R_c/\text{Å}$</th>
<th>$K_S$ / kJ/(mol·nm$^2$)</th>
<th>50</th>
<th>500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.60</td>
<td>0.69</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.61</td>
<td>0.74</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.68</td>
<td>0.75</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

Table S4. RMSIP values for simulations of the closed conformation of LBP with ligand and the open conformation of LBP, quantifying the essential subspace overlap between the AA simulations and the ELNEDIN simulations starting from the equilibrated average AA structure (AAeq). In all cases a force constant for the elastic network of 500 kJ/(mol·nm$^2$) was used, and the cut-off distance, $R_c$, was varied as given in the table. Obtained from the last 80 ns of “real” time of simulation.

<table>
<thead>
<tr>
<th>$R_c/\text{Å}$</th>
<th>Closed</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>0.67</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Table S5. RMSIP values for simulations of the closed conformation of LBP with ligand and the open conformation of LBP, quantifying the essential subspace overlap between the AA simulation and the ELNEDIN or domELNEDIN simulations. In all cases a force constant for the elastic network of 500 kJ/(mol·nm²) was used, and the cut-off distance, $R_c$, was varied as given in the table. Obtained from the last 80 ns of “real” time of simulation.

<table>
<thead>
<tr>
<th>$R_c$/Å</th>
<th>Closed ELNEDIN</th>
<th>domELNEDIN</th>
<th>Open ELNEDIN</th>
<th>domELNEDIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>0.65</td>
<td>0.70</td>
<td>0.74</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Figure S1. The protocol flow for setting up and running simulations using the domELNEDIN approach. Applications and scripts are shown as square boxes with the name in bold and comments or descriptions in plain text. File name examples are given in italic writing.
Figure S2. RMSD as a function of simulation time for ELNEDIN simulations of the closed conformation of LBP with ligand (red line). Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\in \{8, 9, 10\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\in \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S3. RMSD per residue for the ELNEDIN simulations of the closed conformation of LBP with ligand (red line). Obtained from the last 80 ns of “real” time of simulation. Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\in \{8, 9, 10\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\in \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S4. RMSF per residue for the ELNEDIN simulations of the closed conformation of LBP with ligand (red line). Obtained from the last 80 ns of “real” time of simulation. Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\epsilon \{8, 9, 10\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\epsilon \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S5. RMSD as a function of simulation time for the ELNEDIN simulations of the open conformation of LBP without ligand (red line). Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\in \{8, 9, 10\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\in \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S6. RMSD per residue for the ELNEDIN simulations of the open conformation of LBP without ligand (red line). Obtained from the last 80 ns of “real” time of simulation. Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\epsilon \{8, 9, 10\}$ and spring force constant $K_s$ (kJ·mol$^{-1}$·nm$^{-2}$) $\epsilon \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S7. RMSF per residue for the ELNEDIN simulations of the open conformation of LBP without ligand (red line). Obtained from the last 80 ns of “real” time of simulation. Nine different ENs were tested, varying the cut-off distance $R_C$ (Å) $\in \{8, 9, 10\}$ and spring force constant $K_S$ (kJ·mol$^{-1}$·nm$^{-2}$) $\in \{50, 500, 5000\}$. Plots include results from the AA simulation (black line) for comparison.
Figure S8. RMSD as a function of simulation time and RMSF and RMSD per residue for the last 80 ns of simulation obtained from AA (black line) and 80 ns of “real” time of simulation obtained from domELNEDIN simulations (red line) of the closed conformation of LBP with ligand (A) and the open conformation without ligand (B). The force constant parameter, $K_S$, of the EN is in all cases 500 kJ·mol$^{-1}$·nm$^{-2}$ and the cut-off, $R_C$, is either 8 Å or 9 Å as indicated on the figure.
Appendix B to Chapter 2

Protocol Flow

for setting up the domELNEDIN simulation.
Atomistic protein structure in pdb format

**DSSP**
Assign secondary structure elements

**martinize-1.2.py**
with options -elnedyn -ef Ks -eu Rc

**genbox**
CG environment structure info

**generate_domELNEDIN.itp**
CG environment topology info and MARTINI/ELNEDIN FF parameters

**grompp**
GROMACS preprocessor

**mdrun**
GROMACS MD engine

**pDomains** or DomFOLDpdp
Assign protein domain boundaries

**DSSP**

Assign secondary structure elements

**struct.pdb**

**struct.ssd**

**struct.CG.gro**

**domELNEDIN topology**

**struct.domELNEDIN.itp**

**min_md.tpr**

**domdef.dat**

**struct+env.CG.gro**

**genbox**

CG environment structure info

**struct.CG.gro**

**struct+env.CG.gro**

**domELNEDIN topology**

**struct.domELNEDIN.itp**

**genbox**

CG environment structure info

**struct.CG.gro**

**struct+env.CG.gro**

**domELNEDIN topology**

**struct.domELNEDIN.itp**

**grompp**

GROMACS preprocessor

**mdrun**

GROMACS MD engine

1) Program to assign secondary structure
   http://swift.cmbi.ru.nl/gv/dssp/
   Kabsch W, Sander C (1983)
   *Biopolymers* 22, 2577-2637
   Also provided with GROMACS.

2) Server to predict protein domain boundaries
   http://pdomains.sdsc.edu/v2/index.php
   Towards consistent assignment of structural domains in proteins. *J. Mol. Biol.* 339, 647-678

3) Server to predict protein domain boundaries
   http://www.reading.ac.uk/bioinf/DomFOLD/
   PDP: protein domain parser. *Bioinformatics* 19, 429-30

4) Tool to generate CG structure and topology provided by the MARTINI homepage.

5) Standard GROMACS tools.

6) Script developed for the purpose of the present study.