Functional Domain Motions and Processivity in Bacterial Hyaluronate Lyase: A Molecular Dynamics study

Vorgelegt von

Harshad Joshi

aus Pune, Indien

Göttingen, 2007

D7
Referent: Prof. Dr. Tim Salditt
Korreferent: PD Dr. Helmut Grubmüller
Tag der mündlichen Prüfung: 04.05.2007

To my parents and teachers

अल्पानामपि वस्तूनां संहितः कार्यसाधिका ।
तृणैर्गुणत्वापनैर्बध्यन्ते मत्तहस्तिनः ॥

*Great things can be achieved by putting small
things together. Just like a rope made of grass
can be used to tie down a massive mad elephant.
(In the hope that this thesis contributes
in better understanding of the mother nature.)*

Contents

List of Figures

List of Tables

Chapter 0

Prologue

> One's ideas must be as broad as Nature if they are to interpret Nature.
> - Sir Arthur Conan Doyle in "A Study in Scarlet"

In one of his lectures Francis Crick has remarked the following:[1]

> "...The ultimate aim of the modern movement in biology is in fact to explain all biology in terms of physics and chemistry. There is a very good reason for this. Since the revolution in physics in the mid-twenties, we have had a sound theoretical basis for chemistry and the relevant parts of physics. This is not to be so presumptuous as to say that our knowledge is absolutely complete. Nevertheless quantum mechanics, together with our empirical knowledge of chemistry, appears to provide us with a foundation of certainty on which to build biology. In just the same way Newtonian mechanics, even though we know that it is only a first approximation, provides a foundation for, say, mechanical engineering..."

During the last decades modern biology has indeed evolved from just a taxonomical listing to the genomic era. Various attempts have been made to make it possible to solve the complex biological problems with the help of expertise from other branches of science like physics, chemistry or statistics. As a result, branches like biophysics, biochemistry or recently bioinformatics have emerged and a plethora of knowledge from 'live molecules' has been unleashed in front of us. We can 'see' tiny cells or even a single molecule, solve its structural puzzles/determinants or see the relationship with the whole genome

of the species. Modern day molecular biology thus attempts to resolve the mechanism of biological complexity at a molecular level with the help of all these techniques.

X-ray crystallography, Nuclear Magnetic Resonance (NMR) Spectroscopy[2] or modern techniques like x-ray free electron lasers (xFEL), cryoelectron tomography[3] now routinely provide a static picture of bio-molecules for example, the structure of a folded protein. Monitoring dynamical properties of the molecules at the atomic level however, is not trivial and has become possible only recently with advanced experimental techniques like fluorescence spectroscopy,[4,5] single molecule atomic-force microscopy (AFM) combined with either optical or magnetic tweezers,[3,6] NMR relaxation measurements.[7] However, these techniques yet are not able to directly provide an atomistic picture of protein dynamics *in situ* with sufficient time resolution.

Computer simulation techniques have been developed to bridge this gap. Computers have been used to solve scientific problems from as early as the 1950s. With the advent of very powerful computers, numerical simulations have now become a very important tool in research both to bridge the gap between theory and experiments as well as to be a predictive method for the process of interest. A major advantage of computer simulations over the experimental techniques is that it allows to 'see' the process as it happens at atomic detail at any desired time resolution. At the same time, such simulations are obviously only as accurate as the employed molecular methods, and additionally rely on sufficient sampling of the relevant conformational space. There is a number of different approaches in employing molecular simulations as can be seen from Fig. 1, and the type of simulation is usually chosen depending upon the question of interest. Molecular dynamics (MD) simulations are widely used to investigate the processes occurring on timescales from femtosecond to microsecond. It is this technique which we use throughout the present work.

Since the first application of MD to a small protein in vacuum, now more than 25 years ago, advances in computer power, algorithmic developments, and improvements in the accuracy of the used interaction function (force fields, see Sec. 1.1.1) have established MD as an important and predictive technique to study dynamic processes of biomolecules at atomic resolution. Because of the inherent complexity of biomolecules, MD is still limited in time and size scale (Fig. 1). However, if the system under investigation is chosen carefully, MD has shown to be quite powerful to investigate many different processes. The examples include the reversible folding of peptides,[10] or large scale conformational changes in ATPase[11] or chaperones.[12] Such dynamical structural changes are results of the exploration of these molecules in their high-dimensional free energy landscape, and these explorations are often required to fulfil their diverse functions.[13] In terms of such an energy landscape, the kinetics and molecular mechanism as the function requires is thus governed by the energy barriers, valleys and plateaus along the pathways. Since the molecular dynamics approach provides the possibility to probe such energy landscapes, it has become a very important technique for studying structure-function relationship of biomolecules.

We have taken this approach of biomolecular dynamics simulations to probe the
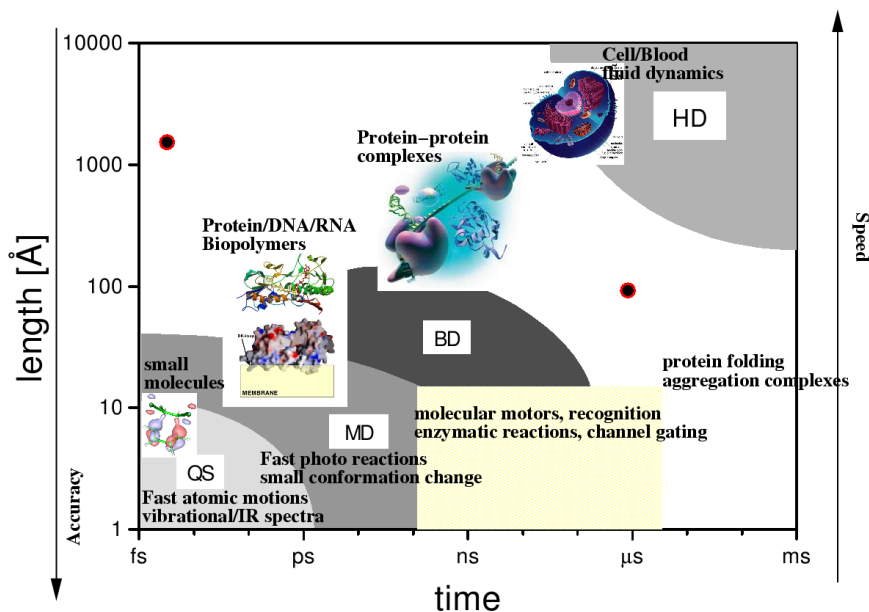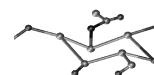
Figure 1: Scheme of the size of the simulation system, computational expense and time scales accessible by different simulation techniques: QS- Quantum simulations, MD- Molecular dynamics, BD- Brownian dynamics and HD-Hydrodynamics/Fluid dynamics. Quantum mechanical methods are accurate but can be applied for only very small systems and for ultrafast processes, while as the system size becomes larger and the timescales of interest increase more and more approximations are needed. The highlighted area near MD (at the ns timescale) shows the timescales and processes accessible by molecular dynamics simulations with todays computer power. The circles show the longest ($\approx 1\mu$s) and the biggest (N > 5x10^9, L$\approx 0.4\mu$m) molecular dynamics simulations (until 2000) by Duan & Kollman[8] and Roth[9] respectively.

atomistic details of a property known as processivity, which is an inherently dynamical property of a certain class of enzymes. In the next two sections, the details on this study and the procedures employed in the present work are outlined briefly.

0.1 Scope of the thesis

As the title already suggests, the main aim of this thesis is to elucidate the molecular mechanism of the processivity by studying a prototypic example of a particular bacterial hyaluronidase from the plethora of other processive enzymes. From the variety of methods of simulations available as discussed before, we deploy the molecular dynamics (MD) approach to tackle this problem at atomistic detail. The question of processivity of bacterial hyaluronidases can be seen from two different perspectives, and we hope to answer both sides in the thesis as it develops. First, the processivity mechanism is observed throughout nature and hence trying to address this issue at the molecular level is of general biophysical/biochemical importance. Secondly, *Streptococcus Pneumoniae*

hyaluronate lyase (*Spn*Hyal), the prototypic enzyme we study here, is itself one of the important virulence factors[14] and has great physiological/medical importance. Thus, by studying this system, we hope to achieve knowledge about the fundamental molecular mechanism in a specific subsystem as well as an overview of the processivity mechanism in general.

Processive enzymes and their biophysical relevance

The processivity of an enzyme can be defined as the number of rounds of catalysis that can be performed before the enzyme dissociates from the substrate.[15] Processivity plays an important role throughout nature and is exploited by many enzymes that synthesise, degrade or modify biopolymers. Processive enzymes presumably enhance their efficiency by remaining attached to their polymeric substrates between multiple rounds of catalysis. Accordingly it is generally assumed that after one round of catalysis, the substrate slides through the enzyme to the next round of catalysis. To date many enzymes are shown to be processive such as a very highly processive enzyme $\lambda-$exonuclease which has a closed toroidal structure that degrades double-stranded DNA (dsDNA) into a single-stranded DNA (ssDNA)[16] or a recently solved phospho-mannomutase/phosphoglucomutase (PMM/PGM),[17] which can utilise either glucose or mannose-based phospho-sugars as substrates. Structural studies have provided first hints on the mechanism of processivity for several proteins.[18] However, atomic details of particularly the substrate sliding phase between subsequent rounds of catalysis, remain largely unknown. The knowledge of this phase would prove important not only to gain structural insight into the molecular mechanism of these processive enzymes but also from a energetic or bioengineering point of view, for example to engineer a processive enzyme mimic.[19]

Molecular mechanism of sugar degradation by hyaluronidases

In the following work, we try to probe the question of processivity mechanism by elucidating the microscopic mechanism of processivity of *Streptococcus Pneumoniae* hyaluronate lyase (*Spn*Hyal), which primarily degrades hyaluronan polymers (HA). Atomistic knowledge of the processive degradation of HA by *Spn*Hyal that may be obtained with the present study has its own fundamental physiological importance. These lyases (Hyals) are an important class of polysaccharide degrading lyases and also have many applications in the medical field. Also, hyaluronan (HA) itself, the polysaccharide that these lyases degrade, has a very important physiological and medical role.

Recently, high-resolution structures of hyaluronate lyase from two different species (*Streptococcus Pneumoniae* and *Streptococcus Agalactaie*) were solved, both of the *apo* form of the enzymes, as well as in complex with substrate and product.[20–23] These structures of HA di–, tetra– and hexasaccharides from the Jedrzejas laboratory are also the first structures of HA bound to biological enzymes. In addition, recently there is also a structure of bee venom hyaluronidase (BVH, currently the best 3D model of mammalian
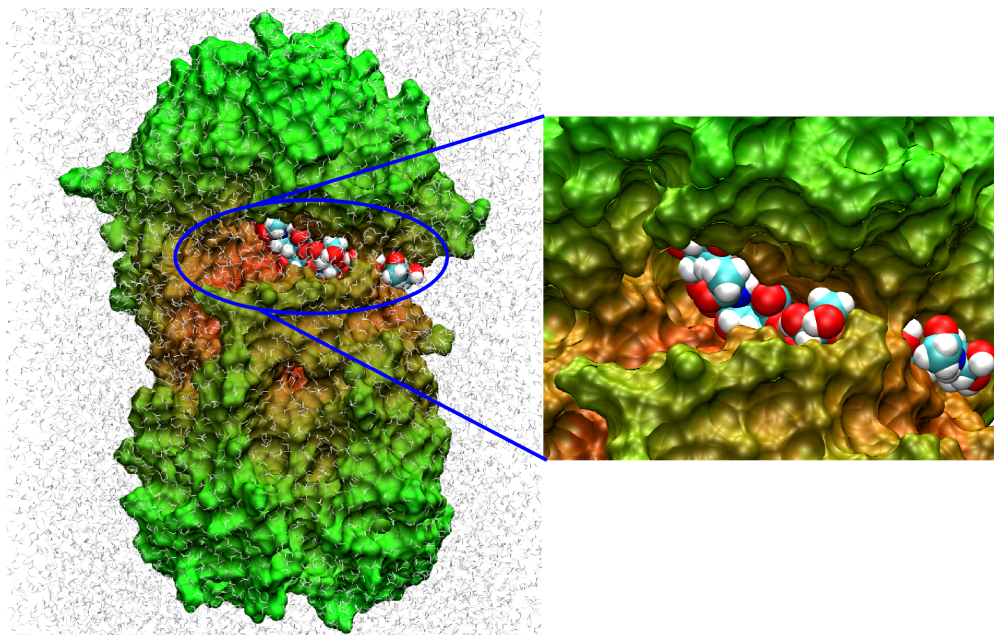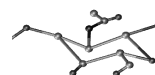
Figure 2: A typical *Spn*Hyal-HA MD simulation system studied in the thesis. Surface representation of *Spn*Hyal with HA inside the cleft (in space filling representation) is shown. The grey lines represent the water molecules as an explicit solvent surrounding the system. The cleft of the protein where the degradation of HA occurs processively has a partially closed shape with room enough for three disaccharide units (or six rings) of the HA. Several free and enforced simulations are performed to investigate the mechanism of translocation of the HA through this cleft.

Hyals) even in complex with HA tetrasaccharide (pdb code: 1FCV (complex with HA) and native enzyme 1FCQ, 1FCU).[24] As these studies only provide state snapshots of the enzyme and enzyme-substrate complex, it is becoming increasingly interesting to know the dynamics of Hyals in their native state as well as in complex with HA substrate. Also it is interesting to compare the structure and dynamics of the HA oligomers under physiological conditions and in complex with Hyal enzymes.

We employ molecular dynamics (MD) simulations with the aim to study one full processive cycle of the *Spn*Hyal-HA system (see Fig. 2), including the substrate translocation by one disaccharide unit. A particular focus is on the question of how the enzyme can provide on one hand strong, specific binding during catalysis and on the other hand weak, unspecific binding during the substrate sliding phase. The project involves a close collaboration with group of Dr. Mark Jedrzejas, an expert in biochemistry and x-ray structural analysis of hyaluronate lyase, who provided experimental validation of and feedback to the simulation results.

0.2   Outline of the thesis
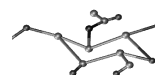
This thesis aims at contributing to the understanding of functional aspects of *Spn*Hyal-HA system at the atomic level. Molecular dynamics simulations of both 'free' and 'enforced' are employed to understand the interplay between the protein dynamics and the sugar translocation. Also the dynamics of hyaluronan oligomers at sub-microsecond timescale is studied. The biological background and the questions for each of the studies is addressed in detail in the introductory section of the corresponding chapters (Chapter 3 through Chapter 6). The thesis is structured as follows:

Chapter 1 — Theory and Methods

Protein dynamics simulations aim at answering specific questions concerning the molecular mechanism and driving forces of the protein transition of interest. They thereby significantly contribute to the understanding of functional motions, a prerequisite for their systematic modification by means of genetic engineering, active compounds, or other means. Molecular dynamics (MD) simulations are widely used to monitor protein motion on up to microsecond time scales nowadays. The method describes the dynamics of a molecular system by numerically integrating the *Newtonian* equations for the respective atoms in the system. The simulations themselves give coordinates and velocities (and energies) of every atom in the system at each timestep and these are analysed afterwards with techniques like Principal Component Analysis (PCA), a method which we use throughout the thesis. The MD method and PCA along with other details are described in Chapter 1.

Chapter 2 — Processivity and Hyaluronidases

Processive enzymes are found throughout nature and many enzymes have been categorised so far as processive enzymes. Chapter 2 gives a brief introduction about what processivity is and what is known about processive enzymes. The prototypic enzyme *Spn*Hyal that we have studied in the present work is also described in detail. The physiological and medical relevance of this enzyme are also sketched. Further insights into the putative mechanism of the processive degradation of the polysaccharide hyaluronan (HA) was obtained by the high-resolution x-ray structures solved in the laboratory of Dr. Jedrzejas. The structures from two different strains showed different structural arrangements giving the first hints that conformational changes of the protein are involved in the mechanism. The structural background of the *Spn*Hyal-HA system and the putative role of protein dynamics as investigated by CONCOORD simulations[22, 23] are sketched in this chapter.

Chapter 3 — Hyaluronan: Structure and Dynamics

The polysaccharide hyaluronan that is processively degraded by *Spn*Hyal has been widely studied in recent years. Due to its high biocompatibility and its common presence in the extracellular matrix of tissues, hyaluronan is gaining popularity as a biomaterial scaffold in tissue engineering research. Its viscoelastic properties primarily make it a very important component in medical use such as in eye surgery (i.e. corneal transplantation, cataract surgery, glaucoma surgery and surgery to repair retinal detachment), treatment in osteoarthritis, cosmetic surgeries and as a tumour marker for prostate and breast cancer. Our studies with HA started primarily with a question of which set of atomic parameters (force field) is appropriate for the simulation studies of this sugar. MD simulations of HA in aqueous solution with different force fields were carried out and the suitable force field was chosen by comparing the simulation data with the available experimental data. These simulations and the choice of the force field are sketched in detail in this chapter.

Chapter 4 & 5 — Flexibility of *Spn*Hyals: Part I & II

The functional roles of proteins are often correlated with their structural changes or their dynamics, and this coupling between the dynamics and function is now a well established theoretical concept.[25] In Chapter 4, we sketch the conformational changes at the active site of the protein that were observed for the first time using a MD approach. The first three principal modes of flexibility of the protein showed drastic differences for the *apo* (in the absence of the sugar) and the *holo* (in the presence of the sugar) configurations. Two newly solved structures were found to confirm this finding, suggesting that the protein actively takes part in processivity and does not just provide the infrastructural framework. In Chapter 5 we sketch the details on several 'free' simulations where we investigated these domain motions in detail with various lengths of HA inside the cleft and at different configurations. All the simulations were investigated for any spontaneous processivity. The structural, conformational and energetic differences between the two sets of simulations, one involving HA in the catalytic phase and the other in the processive phase were investigated to gain insights into the substrate binding and processivity mechanism.

Chapter 6 — Enforced MD simulations of the *Spn*Hyal-HA system

The 'free' simulations from Chapter 4 and 5 indicated that the processivity may occur at longer timescales than previously assumed, since even within $\sim 100\,\mathrm{ns}$ of simulations, a spontaneous processivity was not observed with the translocation of a complete disaccharide unit. This indicated, at first, a strong energy barrier of at least $\sim 20 - 30$ kJ/mol* for the process to be crossed. So, as a next step, we performed enforced or 'biased' simulations on the *Spn*Hyal-HA system in which we forcefully translocated the

---

*when Kramer's rate theory is applied

sugar along an assumed processive pathway. A number of MD simulation techniques have been developed that allow to enforce the system to undergo the transition of interest along a defined pathway.[26–33] Again, the method of choice depends on the questions to be answered regarding the functional motion under investigation. By means of a set of advanced MD simulation methods, this chapter addresses the question of the sugar translocation inside the cleft. First, elevated temperature simulations, force-induced sugar 'pulling', and Essential dynamics simulations were used to enforce and study the translocation phase i.e. the sliding of the substrate through the cleft along the appropriate (putative) reaction coordinate. Secondly, Essential Dynamics and umbrella sampling techniques were used to probe the free energy profile for the sugar on this pathway of the sliding phase. From these studies new insight about the energy barriers for the putative pathway of the sugar in sliding phase has been obtained.

Chapter 7 — Epilogue

Finally in Chapter 7, we summarise all the results from these studies. The general mechanism underlying the processive nature of the enzyme and the interplay between protein and sugar dynamics are discussed.

As stated before, the study of *Spn*Hyal-HA system is a prototype case for processive enzymes in general. It is hoped that the results obtained from these investigations shall add to the bigger picture of processivity mechanism in general.
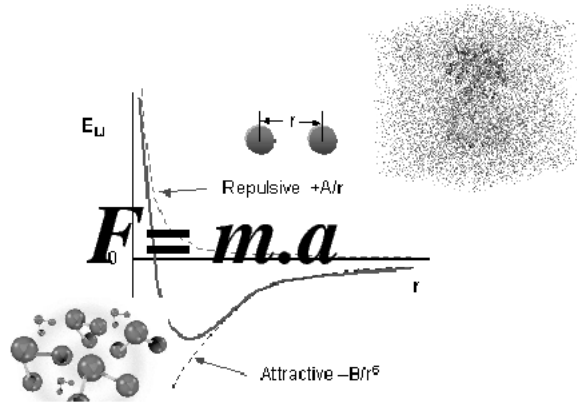
Chapter 1
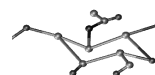
Theory and Methods

All models are false but some models are useful.

- George E. P. Box

Summary

This chapter presents the theoretical basis of the thesis. The details on the particular methods and simulation setup used for different biological systems studied in the present work are given in the respective chapters (Chapter 3 to 6). After a brief introduction, the principles of the molecular dynamics (MD) simulation method, the principal simulation tool used in this work (Sec. 1.1) are discussed. In Sec. 1.2, the conditions for the MD simulations and the methods used in the present work are discussed. Finally, analysis methods such as Principal Component Analysis (PCA) that are used throughout the present work are described in Sec. 1.3.

When it all began...

Newton's *Principia Mathematica*[34] changed the whole perspective of man's quest for understanding the universe, making mathematical physics accepted as a reliable and powerful tool for describing nature. The famous three laws of motion accurately predicted the motions of objects spanning an enormous range of scales — from trajectories of terrestrial projectiles to those of planets, including the legendary windfall apple. These laws worked so well that they were believed to be universal laws for more than two centuries, after which a second revolution in scientific history took place. It was the beginning of the twentieth century, also known as the *'golden period'*,[35] when classical Newtonian mechanics was found to be inadequate to explain the phenomena on the atomic scale, and the theory of *'Quantum Mechanics'*[36,37] was born. Despite the philosophical questions of interpretation[38] which arise from this new theory, it is hard to question the astounding accuracy with which quantum mechanics describes the world around us. Today there is little doubt, that quantum theory applied to electrons and atomic nuclei forms the basis for almost all of physics, chemistry and biology.

Unfortunately, such equations of motion become too complicated to solve analytically for all but the simplest (and hence most trivial) of systems. For describing phenomena of relevance to the larger systems one has to solve these equations approximately by modelling the process of interest computationally. Here again, with the increasing size of the system to be studied, it becomes more and more difficult to obtain the quantum mechanical solution even numerically, and one typically focuses on (sometimes *ad hoc*) approximations to these equations to capture the essential energetics of the problem of interest. Most of the biologically relevant systems fall into this category and hence faster but approximate methods to deal with them have been developed rapidly during the last decades.[39] For example, a systematic reduction of the full quantum mechanical description of a system has been carried out to obtain a more conceptually and computationally managable set of equations which can be applied to large biological systems like proteins. As the method obtained that way forms the working horse used throughout the present work, this method and its principal approximations are discussed in this chapter. We then briefly discuss the basic information that is obtained by solving these equations, but our main focus, to be covered in later sections, is the numerical framework to solve the equations of motions describing the molecular dynamics of proteins, *i.e.* the structure and energetics of proteins at atomic detail as they evolve with time.

1.1 Molecular Dynamics simulations - Principle

This section describes the theoretical framework for the molecular dynamics (MD) simulations. The exhaustive description of the MD-method can be found in recent reviews[40–43] and many textbooks.[39,44–47] Here we just briefly outline the principles and approximations on which the MD simulations are based. As discussed in the introduction, an appropriate simulation technique is generally chosen depending upon the phenomenon of interest. A typical protein consists of tens to hundreds of amino acid

residues, amounting to systems of thousands of atoms. It presents dynamics from bond vibrations at the femtosecond timescale to large-scale conformational changes occurring within microseconds or even slower. The MD simulation method efficiently describes the dynamics of highly complex structures of biomolecules, and yet gives a reasonably accurate description of the phenomena within a large timescale window, spanning sub-picoseconds to microseconds, at atomic detail. In the present work we therefore use MD to study large conformational changes and protein-ligand interactions in the *Spn*Hyal-HA system occurring at sub-microsecond timescale.

1.1.1 From Schrödinger equation to Molecular Dynamics

The observable properties of all forms of matter are determined completely by quantum mechanics from solutions of the time-dependent many-body Schrödinger equation (we stick to the non-relativistic description here),

$$H\psi = i\hbar\frac{\partial\psi}{\partial t},\tag{1.1}$$

with H denoting the Hamiltonian, the sum of potential and kinetic energy; ψ the wave function; and $\hbar = h/2\pi$ with h the Planck's constant. The wave function comprises all the particles of the system, *i. e.* is a function of the coordinates and momenta of both, nuclei and electrons.

This equation includes all the equilibrium properties and nonequilibrium, or response properties. Equilibrium properties encompass all thermodynamic properties, including the equations of state and phase diagrams, and quantities such as specific heat and compressibility. Nonequilibrium properties include responses to various perturbations, such as mechanical impulses, which determine transport and mechanical properties of materials. However, Eqn.1.1, even for small systems of more than approximately ten atoms, is computationally prohibitive with current state of the art computers. To be able to describe larger systems, like a protein in its solvent environment, the following three approximations are necessary.

Approximation I: Born-Oppenheimer

The first approximation is based on the fact, that due to the much lower mass and consequently much higher velocity of electrons as compared to nuclei, electrons can be assumed to instantaneously follow the motions of the nuclei. Therefore, the electronic degrees of freedom can be separated from the nuclear degrees of freedom. This approximation, is called *"Born-Oppenheimer approximation"*.[48–50] Thus, the total wave function ψ can now be separated into the nucleic wave function ψ_N and electronic wave function ψ_e,

$$\psi(R, r) = \psi_N(R)\psi_e(R, r),\tag{1.2}$$
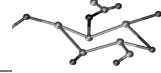
where the electronic wave function depends only on the positions, not on the velocities, of the nuclei; and the resulting time-independent Schrödinger equation for the electrons can then be solved for fixed nuclei positions to obtain the electronic energy,

$$H_e \psi_e(R, r) = E_e(R) \psi_e(R, r) \tag{1.3}$$

where the electron Hamilton operator

$$H_e = T_e + V_{NN} + V_{Ne} + V_{ee} \tag{1.4}$$

is the sum of respective nuclei (N) and electron (e) interactions V, and the kinetic contributions T_e. The fact that the positions of the nuclei enter only parametrically also means that the nuclei now move in an effective potential, given by the ground state energy $E_e(R)$ of Eqn.1.3, which describes the influence of the electron dynamics on the nuclei motion. The motion of the nuclei itself thus follows the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \psi_N(R, t)}{\partial t} = (T_N + E_e(R)) \psi_N(R, t). \tag{1.5}$$

This approximation has been shown to hold to a high level of accuracy.[51–53]

Approximation II: Classical description of nuclear dynamics

Because of their large size, the solution of the nuclear time-dependent Schrödinger equation (Eqn.1.5) is still a prohibitive task, for proteins. As a second approximation, therefore, the dynamics of the nuclei are described by the classical equivalent to it, *i.e.* Eqn.1.5 is replaced by the *Newtonian equations of motion* for classical point masses,

$$
\begin{aligned}
m_i \frac{d^2 R_i(t)}{dt^2} &= -\nabla_i E_e(R_1, \dots, R_N), \qquad i = 1, \dots, N \\
&= -\frac{dV}{dR_i},
\end{aligned}
\tag{1.6}
$$

where R_i and m_i are the coordinates and the mass of atom i and V is the potential energy of the system as a function of all atomic coordinates R_i. Eqn.1.6 is just another form of Newton's second law,

$$F_i = m_i a_i, \tag{1.7}$$

The force F_i acting on atom i at position R_i gives its acceleration a_i. This allows computation of the displacement of atom i within a given time step Δt by integrating the equation to yield a change in velocity, which, in turn yields the displacement. The verlet algorithm[54] or equivalent is typically applied for the efficient numerical integration of Newton's equations of motion.

The classical description mentioned above is appropriate for non-bonded heavy atoms at room temperature. Quantum mechanical effects of nuclei motions, such as the tunnelling of hydrogen atoms[55–57] and the superfluidity of liquid helium,[58,59] become pronounced for light atoms or at low temperatures or also for heavy atoms, at room temperature, when invloving covalent or hydrogen bonds. For this classical approximation to

Type of vibration	$\tau[fs]$
bond stretching	10-24
bond angle bending	20-40
libration of water molecules	50-80
hydrogen bond stretching	170-700

Table 1.1: Typical vibration periods in molecules and hydrogen-bonded liquid. Compare $h/kT = 170\,\text{fs}$ at 300 K.

be valid, the thermal energy, therefore, has to be distinctly larger than the energy gaps between neighboring quantum states of the system. For the harmonic oscillator, as an example, this requirement means that the classical description is valid if $k_B T \gg \Delta E = h\omega$ (k_B: Boltzmann constant; T: temperature; ΔE: energy difference between neighbored states; h: Planck constant; ω: vibrational frequency). At physiological temperatures (300 K), this condition is fulfilled for vibration periods $\tau \gg 160\,fs$. From Table 1.1 it can be seen that, bond stretching and bond angle bending modes, as well as librations of water molecules* involve shorter vibration periods and, hence, are beyond the classical limits.[60] The conformational motions of interest in the present work, however, occur on picosecond[61] or longer timescales and can therefore be described well within classical mechanics framework.

Using the Born-Oppenheimer approximation and a classical description of nuclei dynamics allows to describe systems of several hundred atoms on timescales of about $10\,\text{ps}$.[62] The large number of electrons in proteins still prohibits to solve the time-independent Schrödinger equation for the electrons. Therefore studies like the present one require a further approximation, in which the potential energy of the whole system is calculated classically, on the basis of a simple 'empirical' model for atoms connected via bonds.

Approximation III: Molecular Mechanics

In the Born-Oppenheimer approximation the nucleic coordinates enter only as parameters and move in an effective potential $E_e(R_1, \ldots, R_N)$ of Eqn.1.6. In studies as the present one, where no chemical reactions are involved, this effective potential can therefore be approximated by a semi-empirical function, or *"force field"*. The development of a force field is a major field of computational chemistry.[63] Here we just outline the principles behind it. The new potential energy as a function of the atomic coordinates approximately describes all bonded and non-bonded interactions. In bonded interactions covalent forces caused by the change in length of a chemical bond, V_b (Fig. 1.1 A); the bending of an angle between two bonds V_a(B); the out-of-plane deflection of planar

---

*rotational vibrations whose frequency is determined by the moment of inertia of a water molecule and forces working on this mode, which are mainly due to hydrogen bonds.
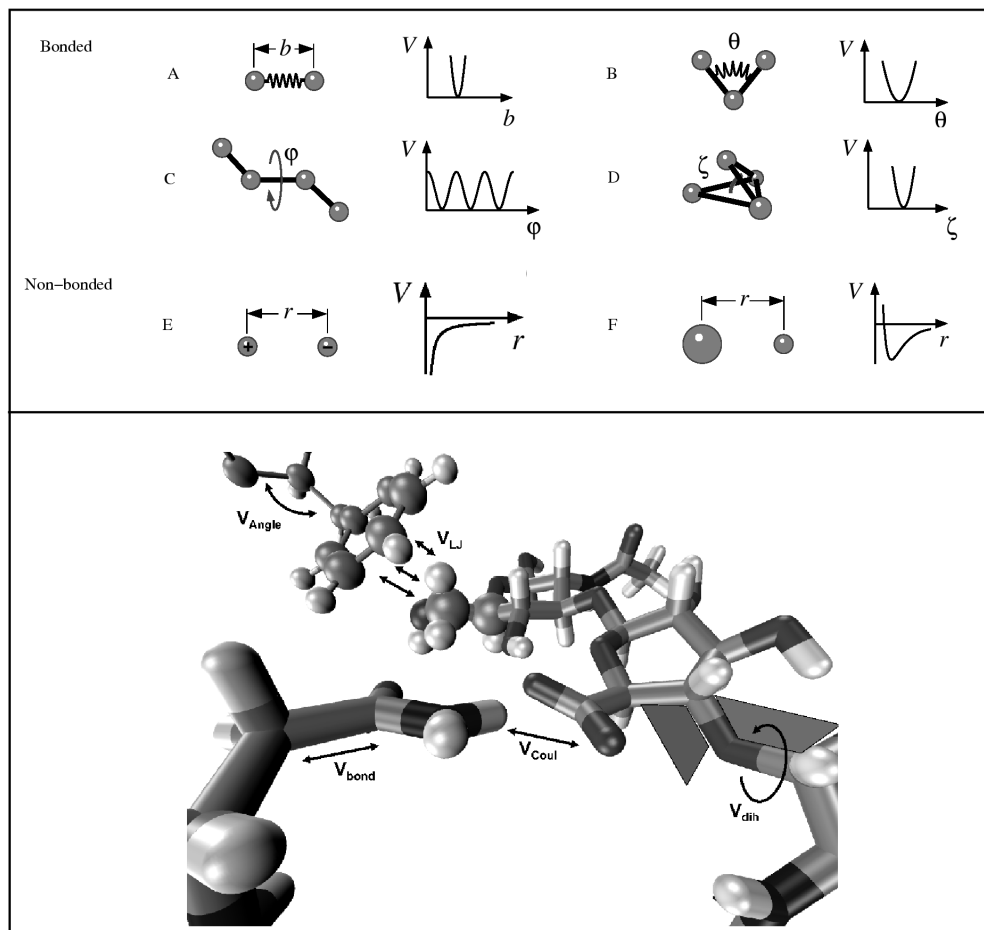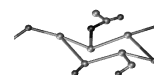
Figure 1.1: Scheme to illustrate the force field terms. The forms of the potentials used for each of the terms in a typical mechanical force field are shown in the upper panel. The lower panel shows these interactions by means of the example of the HA with protein side chains. Black arrows show the corresponding bonded and non-bonded interactions - A: V_{bond}, bond-stretching potential; B: V_{Angle}, angle-bending potential; C: V_{dih} dihedral (out-of-plane) potential. E: V_{coul}; Coulomb potential, F: V_{LJ}, Lennard-Jones potential.

groups, V_{imp} (C); and the torsion of bonds (D),V_{dih}; are described, and the associated potentials are chosen as harmonic(A-C) or periodic (D), respectively. Non-covalent interactions are: non-polar *van der Waals* interactions V_{vdw}(E) and the electrostatic interactions between partial and formal charges of the atoms, V_{coul} (F). In *van der Waals* interactions, a Lennard-Jones[64,65] potential (E) collectively models the short-range repulsion (preventing atoms from penetrating each other) and induced dipole attraction (dispersion) between two atoms. The total potential is the sum of all these individual interaction terms, and is given by
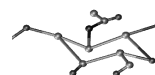
$$
\begin{aligned}
V(R) \;=\;& V_{\mathrm{b}} + V_{\mathrm{a}} + V_{\mathrm{imp}} + V_{\mathrm{dih}} + V_{\mathrm{vdw}} + V_{\mathrm{coul}} \\
=\;& \sum_{\mathrm{bonds}} \frac{k_i}{2}\left(b_i - b_{i,0}\right)^2 \\
&+ \sum_{\mathrm{angles}} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{\substack{extraplanar \\ angles}} \frac{k_i}{2}\left(\zeta_i - \zeta_{i,0}\right)^2 \\
&+ \sum_{\mathrm{dihedrals}} \frac{V_n}{2}\left(1 + cos\left(n\phi - \delta\right)\right) \\
&+ \sum_{\substack{atoms \\ i}} \sum_{\substack{atoms \\ j>i}} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{4\pi r_{ij}},
\end{aligned}
\tag{1.8}
$$

Here the force field parameters for bonded interactions are the equilibrium bond length $b_{i,0}$ and angle $\theta_{i,0}$, the respective force constants k_i, and the multiplicity n, the energy height V_n and the phase δ of the dihedral. Non-bonded interactions are parametrised in terms of partial charges q_i for Coulombic interactions, and the parameters ϵ_{ij} and σ_{ij}, defining the depth and position of the Lennard-Jones potential. These parameters are derived by fitting data to experimental thermodynamic or structural data such as crystal structure, infrared spectra, free energies of solvation etc., combined with high level *ab initio* (quantum) calculations in a self-consistent manner.[43, 44] In the case of proteins, such parametrisation is typically carried out for small simple molecules, like amino acid analogs and short peptides, which can be regarded as the elementary building blocks of large proteins.

A series of different force fields have been developed (CHARMM,[66] GROMOS,[67] AMBER,[68] OPLS,[69, 70] JUMNA,[71] SPASIBA,[72] CFF,[73] PFF01,[74] MM3[75] etc.), which differ in the type of described molecules for which they are optimised, the number of considered degrees of freedom, the methodology of optimisation and the exact formulation of some of the energy terms in Eqn. 1.8. In the present work, the OPLS-AA[69, 70] force field is used for both the protein Hyal and the polysaccharide HA, with a small modification in the HA parameters[76] (see also Appendix B for the complete set of parameters file and Sec. 3.2 of Chapter 3 for the details on the parameter choice of the sugar). The OPLS-AA force field has been developed and refined for proteins by quantum mechanical calculations of small chemical entities, and treats all atoms explicitly. It has become one of the popular force fields for the simulation of protein dynamics and is under constant development.[77] It has been applied here with the TIP4P water model* in conjuction with which this force field has been optimised.[78]

In the present work, the applied force field describes the potential energy of proteins for the simulation of protein dynamics. Since molecular mechanical force fields do not

---

*A great web-resource on the water structure and models can be found at: http://www.lsbu.ac.uk/water/

consider the change of the electronic wave function with respect to the nuclei, they do not treat quantum effects, such as polarisation, chemical reactions, or electronic excited states, explicitly. Nevertheless, the modern force fields implicitly incorporate quantum effects like polarisation and excited state properties to some extent.[79] This demonstrates the usefulness, and, at the same time, the limit of today's empirical force fields.[80] It is thus crucially important to validate and complement MD simulations by experiments wherever possible, as is also consistently carried out throughout this thesis.

1.2 MD simulations in practice

In the previous section we saw how the exact quantum mechanical description of large biological systems can be reduced to the set of force-field based Newtonian equations of motion. All the simulations carried out in this work were performed using such a framework as implemented in the GROMACS software package.[81–83] The algorithms and other methods used by this software in a typical MD simulation study and the ones used in the present work will be introduced in the following subsections: In Sec. 1.2.1 the *leap-frog* algorithm,[84–86] an efficient algorithm used for the numerical integration of the equations of motion (Eqn.1.6) is described. In Sec. 1.2.2 various ways to account for the solvent environment are discussed and the choice of the atomistic model is justified. The chosen boundary conditions and the conditions to generate an isothermal/isobaric ensemble are discussed in Sec. 1.2.3 and Sec. 1.2.4 respectively. Sec. 1.2.5 sketches some of the techniques used to improve the efficiency of MD calculations. Finally, MD simulations usually imply a minimisation and equilibration phase, the relevance of which is explained in Sec. 1.2.6.

1.2.1 Integration method and time step

A prerequisite for describing the dynamics of a molecule by means of the Newtonian equations of motion is a proper initial structure and initial atomic velocities. Structures with an atomic resolution are usually obtained from x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.[3] Protein structures can be found in the Brookhaven Protein Data Bank,[87,88] which contains, up to now, more than $40,000$ structures. Small simple molecules can be built using molecule editors (QUANTA, insightII,[89] Molmol[90] etc.). The initial atomic velocities are generated at the start of the MD simulation, Maxwell distributed according to the appropriate temperature.

Based on these initial conditions, Newton's equations of motion (Eqn.1.6) are numerically solved iteratively in small time steps, Δt. There are many algorithms present (like Verlet,[54] Verlet-Störmer,[91] *velocity-verlet*,[92] Beeman,[93] Runge-Kutta,[94,95] leap-frog[84–86]) to perform this integration (For more detailed discussions about the integrator methods see[45,47,96–98]). In the present work, the numerically stable *leap-frog**[* algorithm

---

*Notably, an interesting description of the leap-frog algorithm can be found in *"The Feynman Lectures on Physics"*, Vol.1, Chapter 9. see ref.[99]

is employed,[60] which is of the second order in Δt and the advantage of this method is that the expensive force calculation is done only once per integration step. The algorithm calculates positions $\mathbf{r}$ at time t and velocities $\mathbf{v}$ at time $t - \frac{\Delta t}{2}$,

$$
\begin{aligned}
\mathbf{v}\left(t + \frac{\Delta t}{2}\right) &= \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}(t)}{m}\Delta t \\
\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \, .
\end{aligned} \tag{1.9}
$$

The integration time step, Δt in Eqn.1.9 has to be chosen such that it is small (typically about a factor of ten smaller) in comparison to the fastest motions of the systems to ensure stable and accurate integration. As mentioned in Sec. 1.1.1, since hydrogen-bond vibrations occur within several femtoseconds (Table 1.1) this restricts the time step to $\sim 1\,\mathrm{fs}$. To perform faster calculations for large system larger time steps can be used by using algorithms that constrain bond lengths, (see Sec. 1.2.5). MD simulations presented in this work apply the LINCS algorithm, allowing a time step of $2\,\mathrm{fs}$.[100] The discretised equations of motion (Eqn.1.9) are modified further to model the effect of a heat bath and to treat the system in an isothermal (NVT) or isobaric (NPT) ensemble.

1.2.2 Solvent environment

The solvent environment strongly affects the structure and dynamics — and hence the function — of proteins and bio-polymers[101, 102] and therefore must be described accurately. In the present work, the solvent environment of the protein was simulated explicitly (see Fig. 1.2). An explicit solvent environment, however demands high computational costs, requiring up to 90–95% of the total simulation time. To alleviate this problem, many mean field models for treating solvent effects implicitly have been developed.[103–106] Although they save computational power, these treatments suffer from severe limitations as discussed below. We will also discuss why, in the present work, the explicit description of solvent molecules is indispensable.

Implicit solvent model

Solvent molecules influence the properties of the solute in many different ways. First, polar solvent molecules force the solute, say protein, to minimise its hydrophobic surface, which may strongly affect the conformation and stability of the solute. This *hydrophobic effect*[107] can be approximated in implicit solvent models by introducing a hydrophobic surface dependent energy term (proportional to solvent accessible surface area of solute atoms or groups of atoms[103, 104, 108–111]) to the force field.[112] Furthermore, the dielectric shielding of molecular charges due to the polarisability of the solvent can be roughly described by using a distance dependent dielectric coefficient $\varepsilon_r > 1$ ($\varepsilon_{\mathrm{water}} \approx 80$ and $\varepsilon_{\mathrm{methanol}} \approx 30$).

However, implicit models impose another level of approximation to the method and may lead to wrong conformations of simulated peptides or proteins. To name an example,

a model using surface area terms as well as a dielectric function[113] appears to favour helices with backbone hydrogen bonds between residues i and $i+5$, the — extremely rare — π-helices, over those with bonds between residues i and $i+4$, the – very common – α-helices.[113,114]

Also, at physiological concentrations, an additional dielectric screening arises because of ions. A much more computationally expensive method to describe this dielectric shielding is to solve the linearised Poisson-Boltzmann (PB) equation at each point of the simulation system.[104,115–117] The computational expense spent on solving the PB equation has been reduced by updating the potential only every 200 integration steps[115] instead of at every step. However, comparisons of electrostatic potentials calculated by a finite-difference PB approach, and from MD simulations with an explicit solvent environment shows a difference of at least 0.4 $kcal\ mol^{-1}\ e^{-1}$, demonstrating the limited accuracy of the implicit solvent model.[118] Apart from these electrostatic effects, the dynamics of the solute is influenced by the viscosity of the solvent. Langevin dynamics[119,120] principally allows to describe this effect by introducing noise and friction forces.[121–126] However, the translational and rotational diffusion depends on the particular interactions between the solvent and the solute. Since in this work the translational diffusion of a polysaccharide ligand molecule attached to a protein shall be studied in detail, the ligand-solvent interactions and protein-solvent interactions must be described as accurately as possible. Likewise, specific interactions like H-bonds between protein, sugar and water are important to be described explicitly. Moreover, the solvent properties in the vicinity of a protein, like, e. g., the viscosity might significantly differ from its bulk properties.[127–129] Therefore, the usage of explicit solvent molecules in this work is mandatory.

Chosen solvent models

For most proteins, except membrane proteins, the natural environment is water. Therefore, to be able to mimic the system under study as close to the *in vivo* system as possible, generally proteins to be simulated are solvated in water with ions in physiological concentrations. In the present study, we have also chosen the fully explicit water solvent environment, with TIP4P water model,[130] for all our simulations. Also in all the simulations Na^+ ions were added to neutralise the net charge of the protein-ligand system (counterions).

1.2.3 System boundaries

Generally, a system in the simulation set up is many orders of magnitude smaller than in the respective experiment, to keep the simulation computationally tractable. Even if explicit solvent conditions are chosen, another major difference between the molecular dynamics simulations of a system and the experiments performed on the same system are the boundary conditions. To minimise artefacts arising from the finite-system boundaries, which become important for such small system sizes, appropriate boundary

conditions have to be chosen. Different solutions have been suggested like (1) introducing a boundary potential that maintains the volume and shape of the simulations system or counterbalances the surface tension (2) preventing the solvent molecules from evaporation by a 'wall' of atoms which are fixed[131] or harmonically restrained at stationary points (3) allowing stochastic dynamics for the atoms close to the wall to account for the energy exchange with the environment and (4) avoiding preferred orientations of the solvent molecules on the surface.[39, 131–135] In the present work, *periodic boundary conditions* (PBC) are employed to overcome any short-range surface artefacts. These also facilitate pressure coupling (Sec. 1.2.4). In PBC, each simulation box is periodically replicated an infinite number of times so that each copy is surrounded by virtual copies of itself in all directions. Possible shapes of the unit cell can be a cuboid, a dodecahedron, or a truncated octahedron. In a box with periodic boundaries, a molecule that leaves the box on one side, immediately reenters the box on the opposite side. In this way, the simulation system does not have any surface. However, artefacts may arise from this artificial periodicity, since the molecules also interact with their periodic images due to the long-range electrostatic interactions. If these interactions are strong enough, they may even strongly affect the conformational dynamics such as the artificial stabilisation of α-helices.[136] These artefacts are minimised by increasing the box to an extent large enough such that the system of interest (say protein) does not interact with its image from the neighboring box. The most popular way used is to apply the *minimal-image convention*, where each atom 'sees' at most only one image of every other atom in the system. The energy and/or force is calculated with the closest atom or image.

In the present work, a cuboid box was chosen for all simulations. The choice of the box size is a trade off between minimisation of artefacts from the periodic boundaries and maximisation of the computational efficiency. For the studies of HA oligomers the starting condition of the box was such that the fully extended HA is buried inside the box at least 1 nm away from the box walls (see Fig. 1.2 a). For all the Hyal-HA systems, the minimal distance between protein and the borders of the wall was at least 1 nm (see Fig. 1.2 b). For detailed discussion about the particular simulation set up see the respective sections.

1.2.4 Temperature and pressure coupling

In principle, the molecular dynamics calculated by solving the Newtonian equations of motion (Eqn. 1.6) conserves the total energy of the system (NVE ensemble) and depends on the initial conditions only, such that after a long time, the ensemble of structures of the respective trajectory will approach a microcanonical ensemble*. In real systems, however, a molecular subsystem (say protein) of the size studied in the simulation constantly exchanges energy with its surrounding, making the temperature, rather than energy, remain constant (canonical ensemble). In addition, numerically

---

*assuming that a trajectory of infinite length will pass every phase space point with its appropriate Boltzmann probability (ergodicity), or will pass arbitrarily close to every such point (quasi-ergodicity). For detailed discussion refer to the standard textbooks on statistical mechanics.[120, 137, 138]
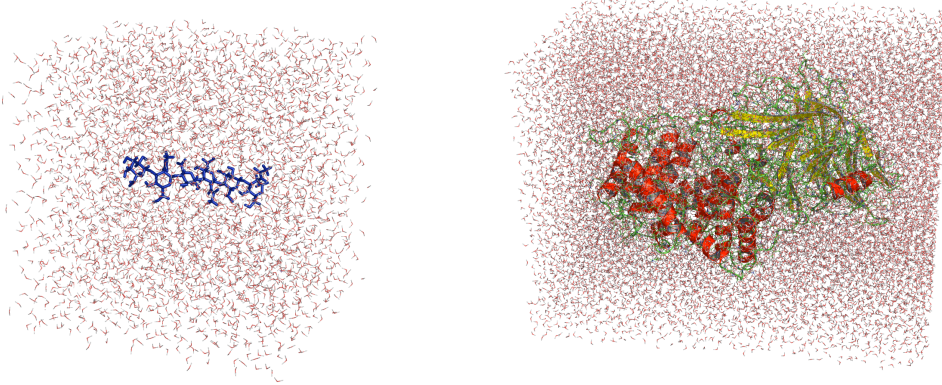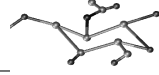
Figure 1.2: (a) Setup for simulations to study the hyaluronan oligomers. Hyaluronan hexas-accharide is shown in stick representation (blue). (b) Simulation setup for the Hyal/Hyal-HA system. The protein is shown in cartoon representation. Water molecules are shown in line representation in both pictures

solving the Newtonian equations leads to discretisation errors[95,97] and rounding of the force terms introduces numerical noise, i. e., random forces, which heat up the system. To be closer to reality, energy exchange should therefore be introduced to the simulation.

It is thus necessary to control the temperature T of the system, which should remain close to a given target temperature T_0. Several methods to achieve this by coupling the system to a heat bath have been proposed.[139–141] The *Nosé-Hoover thermostat*,[139,140] e.g., induces a heat bath parameter as a new degree of freedom whose dynamics is driven by the deviation of the real system temperature from the heat bath temperature; here a respective 'inertia' parameter governs the strength of the coupling. This approach has the disadvantage that it produces a spurious oscillatory relaxation.

In the present work, therefore, the *weak coupling scheme* of the *Berendsen thermostat*[141] is used, where the coupling to a heat bath is achieved by correcting the actual temperature according to

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad , \tag{1.10}$$

which leads to a strongly damped exponential relaxation of the temperature towards the target temperature T_0 with a time constant τ. The change of the temperature is achieved by rescaling the velocities of each atom every step with a time-dependent factor λ given by

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)} \quad . \tag{1.11}$$

The time constant τ depends on the parameter τ_T

$$\tau = \frac{2C_V \tau_T}{N_f k_B} \quad , \tag{1.12}$$

where N_f denotes the total number of degrees of freedom, C_V the heat capacity, and k_B Boltzmann's constant. In all explicit solvent simulations presented in this work, the target temperature T_0 and the coupling time constant τ were chosen to 300 K and 0.1 ps respectively, unless otherwise specified.

In addition to the heat bath coupling, real biological systems are subjected to a constant pressure of usually 1 atm. Therefore, again to mimic the experimental situation as closely as possible, in the simulations, isobaric ensembles were generated, *i.e*, the pressure, rather than the volume, was kept constant. In the present work, this is achieved by using a similar approach as for the temperature coupling. Now the pressure is corrected in each step (*Berendsen barostat*[141]) according to

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \qquad (1.13)$$

The pressure thus relaxes exponentially to the target pressure P_0, which was chosen to 1 atm in all simulations. The pressure correction is achieved by scaling the coordinates, at every integration step, by a factor μ, given by

$$\mu = 1 - \frac{\Delta t}{3\tau_p}\kappa\left(P_0 - P\right), \qquad (1.14)$$

where κ is the isothermal compressibility of the system. In the present work, the value for water at 1 atm and 300 K, $\kappa = 4.6\ e^{-5}bar^{-1}$, was used. The pressure P was calculated from the kinetic energy E_{kin} and the virial coefficient Ξ according to

$$P = \frac{2}{3V}\left(E_{kin} - \Xi\right), \qquad (1.15)$$

where V denotes the volume of the system. The virial Ξ is defined as

$$\Xi = -\frac{1}{2}\sum_{i<j}(\mathbf{r_i} - \mathbf{r_j}).\mathbf{F_{ij}}. \qquad (1.16)$$

Here, $\mathbf{F_{ij}}$ denotes the force exerted from atom i on atom j. For all the simulations in the present work, all directions were scaled isotropically. Due to large fluctuations of the instantaneous pressure in such small systems, a weak coupling constant of $\tau_p = 1\,\mathrm{ps}$ was applied.

1.2.5   Improving efficiency

For large proteins such as studied in the present work, molecular dynamics simulations described above are still very highly demanding in computational expenses and hence optimisations are routinely applied to improve the efficiency of the simulations[*]. These

---

[*]A 20 ns MD simulation for a typical explicit solvent system of $50,000$ atoms with exact calculation of non-bonded forces (*i.e.* $O(N^2)$ complexity) would require about $6 \cdot 10^{16}$ floating point operations (flops). Assuming a typical (actual) computing power of 1.0 Gflops/s this calculation would take about 40 years. The efficient methods used in GROMACS and in this thesis (larger timestep, PME electrostatics *etc.*) make this calculation possible in only a few months. Furthermore, parallelisation means that for the perfect(ideal) scaling, the time decreases by a factor of two for every doubling of the computer resources.

can be a larger time step, using compound atoms or even course grained models, applying efficient methods to calculate the non-bonded force terms in (Eqn.1.8), and by using many processors at a time (parallelisation).

Larger integration timestep

Bond-length constraints do not significantly alter the protein dynamics,[142] and when excluding the bond-stretching vibrations, the HOH bond angle bending becomes the fastest degree of freedom, with a period of 21 fs. (See Table 1.1). Therefore, as discussed in (Sec. 1.2.1), constraining the bond lengths allows an integration time step Δt (in Eqn.1.9) of 2 fs without any loss of generality. If angle constraints are also applied, the vibrations below 50 fs can be removed (See Table 1.1), thus allowing an even larger integration time step of 5 fs. However, such constraints have been shown to affect the conformational dynamics.[142] For these reasons, in the present work, a timestep of 2 fs was used with the bond-length constraint algorithm LINCS,[100] which after an unconstrained integration step, rescales the bond lengths to their target value. For water, the SETTLE algorithm was used which incorporates the constraints within the potential by means of Lagrange multipliers.[143]

Efficient calculation of the non-bonded forces

In each integration step, the force calculation, particularly the evaluation of the non-bonded interaction is computationally the most expensive part. The non-bonded force terms are the *van der Waals* force (Fig. 1.1 E) and Coulomb potential (Fig. 1.1 F), and calculation of these forces requires all interatomic pairs to be considered (last term of Eqn.1.8). Thus the complexity of the calculation scales as $O(N^2)$ for an N atom system. In a simple and routinely used way to enhance the computational efficiency, the calculation of non-bonded interactions is performed using a cut-off scheme.[66] In such a cut-off approach the interactions are calculated only for the atoms residing within a certain defined cut-off (typically in the range of 1.0 to 1.4 nm). However, this approach is a good approximation only for the short ranged *van der Waals* potential. For the long ranged Coulomb forces, a twin-range-approach[45] has been used in the past where the forces are additionally, but less frequently calculated for atoms within a larger shell, where they fluctuate more slowly than closer atoms. However, the use of such methods still creates artefacts[144, 145] of truncating interactions suddenly beyond a certain range. To calculate the electrostatic forces, we therefore applied the grid-based Ewald summation method* — called Particle-Mesh-Ewald (PME)[147, 148] — which is more accurate albeit computationally more expensive. This method, as implemented in GROMACS, uses a grid ('charge-grid') to assign charges and fast fourier transformations for the calculation of the reciprocal sum, and therefore scales only as $O(N \cdot logN)$ as opposed to $O(N^2)$ scaling of the ordinary Ewald method. In the presented work, the PME nethod was used with

---

*The Ewald method was first introduced to calculate the long range interactions of the periodic images in crystals,[146] and is valid in the present work because of the use of periodic boundary conditions.

a neighbour-list update, Coulomb and *van der Waals* cut-off $r_1 = r_2 = 10\,\text{Å}$, and the charges were assigned to a grid with a lattice constant of $1.2\,\text{Å}$ using cubic interpolation. This means that electrostatic interactions within $10\,\text{Å}$ are calculated explicitly, whereas long range elctrostatic interactions are interpolated in the grid. Both the explicit and PME interactions are updated every step.

Parallelisation

For a large biomolecular system such as described in the present work, the molecular dynamics run can still be remarkably slow for the long time-scale phenomena. The run time of the simulations can be shortened significantly by running each simulation on more than one processor in parallel. Here, the atoms of the simulation system are distributed to different processors such that the workload of each processor is maximised while keeping the required data transfer at minimum. The GROMACS package is very well parallelised, and we used 2, 4 or 8 processors for the different simulations.

1.2.6 Minimisation and equilibration

The simulation systems are generally relaxed to an equilibrium for a given temperature prior to the production runs (the simulations from which the analyses are made). This relaxation is carried out in two steps namely, energy minimisation and equilibration, in which the kinetic energy is equally distributed to all degrees of freedom. This relaxation is necessary, for several reasons.

Molecular dynamics simulations are generally started from a well resolved x-ray or NMR structure. For the present work the x-ray structure of the resolution $1.7\,\text{Å}$ was used.[22] These structures — or the coordinates of the atoms in the molecules — are determined from the electron density maps[2,3] and then refined by procedures like simulated annealing and minimisation, in which the knowledge about covalent geometries enters.[39,149,150] Nevertheless, due to the trade-off between force field and goodness-of-fit to electron density maps, the resulting structures can show abnormal bond lengths, considerable deformation of bond or torsion angles, and even overlap of *van der Waals* spheres. Additionally, some internal strain may also be induced by transferring a protein from a crystal environment (including crystal packing interactions) to the solvent environment. Such strains can lead to high forces, and if started from such a structure, the numerical simulation may fail (resulting in a "crash") due to integration artefacts occurring due to such high forces. Thus, a step before MD is required to avoid such unfavourable forces. In the present work, these strains were relaxed by a steepest descent on the energy landscape, until a local minimum was reached. In this iterative procedure an initial step size of $0.1\,\text{Å}$ was chosen, adapted to the steepness of the potential, and the descent was continued until the step size converged to the computer precision. Subsequent to the minimisation, the system is solvated with water and counterions and then coupled to a heat bath (Sec. 1.2.4) and heated to the target temperature $300\,\text{K}$ (unless otherwise stated) and then equilibrated at this temperature. In the present work, this is

achieved by restraining the atomic positions in the simulation for 2 ns. In the position restrained simulations, the heavy atoms of the system were kept close to their initial positions $\mathbf{r_0}$ by subjecting them to a harmonic potential.[60]

$$V(r) = \frac{k_r}{2}(\mathbf{r} - \mathbf{r_0})^2 \tag{1.17}$$

After this equilibration step the solvent environment (water and counterions) is equilibrated or "relaxed" around the protein or protein-ligand complex, while making sure that the experimentally determined protein structure is preserved. The system is then subjected to further equilibration with harmonic constraints released. Whether the system is equilibrated or not is monitored by plotting the relevant observables (e.g. root mean square deviation and potential energy) versus time. The equilibration is considered to be reached when drifts of these observables are absent or small. The trajectories succeeding the minimisation and equilibration phase, were taken for the calculation of thermodynamical properties discussed in Sec. 1.3 or further analysis.

1.3 Analysis methods

After equilibration the trajectories can be analysed for various equilibrium and non-equilibrium properties. In the simulations presented in this work, the configurations of the system after every *picosecond* were recorded for further analysis. This section describes one of the most used analysis methods in the thesis, namely Principal Component Analysis (PCA) (Sec. 1.3.1).

1.3.1 Principal Component Analysis

In a complex energy landscape of a protein, it is seen that rare transitions between multiple minima are often also the functional motions. Hence methods have been under constant development to probe such areas of the landscape (see Chapter 6). Analysing an MD simulation of a protein for such functional motions is not a trivial problem. It requires to filter out all the fast motions of atomic wigglings and extract the collective or functional motions occurring on slower time-scales.[151] Currently, the two most widely used methods to determine such motions are: normal mode analysis (NMA)[39,152,153] and principal component analysis (PCA).[151,154–156] The first attempts to filter such collective degrees of freedom in a protein dynamics were realized by normal mode analysis (NMA) of small proteins[157–160] where the potential energy is approximated harmonically and the collective modes are obtained by diagonalizing the Hessian matrix in a local energy minimum. Apart from small molecules, the NMA approach, has also been successfully applied to calculate high frequency vibrational spectra of proteins.[161] However, it is not clear to what extent this harmonic approximation to a single minimum of the potential energy surface can characterise the functional motion on the complex frustrated multi-minima energy landscape of proteins.[162]

17

Principal Component Analysis (PCA) has been in use for various different applications including data analysis, data compression, data visualisation, image processing, pattern recognition and has proved to be a very useful application for analysing high-dimensional data in a relevant reduced-dimensional subspace. Since its first application to macromolecules,[163] PCA has also been one of the most successful methods for analysing protein dynamics.[164] PCA and other quasi-harmonic analysis methods[165–167] show that beyond the harmonic approximation protein dynamics is dominated by few, mostly anharmonic, collective modes, describing about 90% of the total atomic fluctuations in a protein with only first few (5-10 %) collective degrees of freedom.[155]
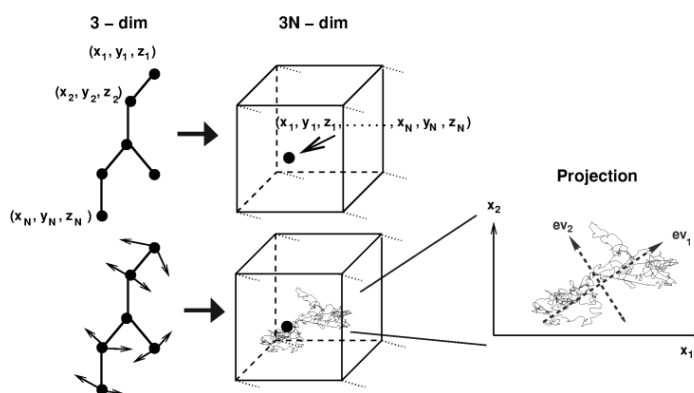
Figure 1.3:   A schematic representation of PCA for an MD trajectory.
source: http://www.mpibpc.gwdg.de/abteilungen/073/presentations/Overview/pca.html

PCA is based on the principle of diagonalisation of the covariance matrix of the atomic fluctuations calculated from an ensemble of structures giving the eigenvectors as the directions of the largest fluctuations. The eigenvalues correspond to the amplitudes in those directions. From a PCA of protein dynamics we get two kinds of motions. The first type constitutes the anharmonic motions caused by rare transitions between multiple minima. These usually describe large (domain) motions occurring on long timescales. They are often referred to as 'essential' modes of motion as they frequently have been shown to be functionally relevant. The second type constitutes so called quasi-harmonic modes, since the fluctuational distribution of these modes can well be approximated by a Gaussian. The 'essential dynamics' or the dynamics in the essential subspace is often most interesting and at the focus of computational studies.[12,168] Enhanced sampling techniques (see Chapter 6) or simple models of protein dynamics[169,170] have been developed based on this notion. There have been different opinions about the convergence of PCA modes in a simulation[171,172] and whether the PCA modes obtained from a short MD simulation really describes a considerable and sufficient amount of the protein motion observed on long time scales. However, these opinions argue about the PCA from short MD simulations while most of the simulations presented in this work are on sub-microsecond timescales. We show that for those simulations the principal modes are not only sufficiently converged but also favourably agree to experiments. The principle of PCA is illustrated in Fig. 1.3

Pneumolysin

PspA
Hyaluronate lyase
CbpA
Autolysin
Neuraminidase
PspA
PspC

Cell membrane   Cell wall

Cleft
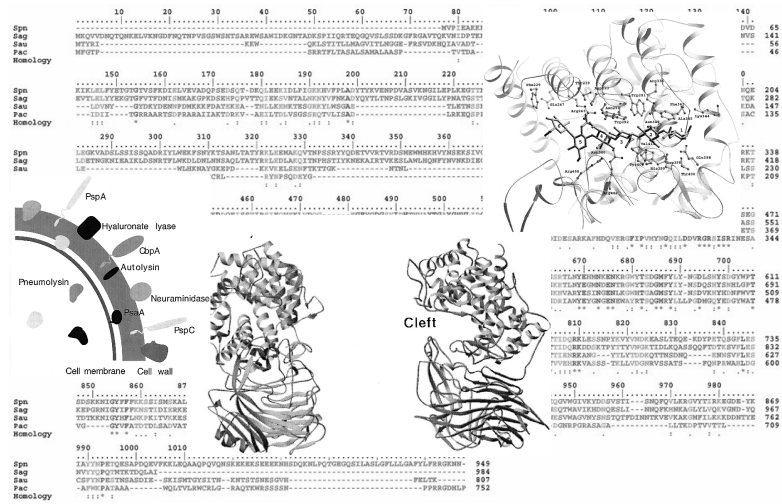
Chapter 2
Processivity and Hyaluronidases

Summary

Processive enzymes are a special class of enzymes that can remain attached to their polymeric substrates between multiple rounds of catalysis. Because of this property the efficiency of these enzymes is increased manifold as compared to the generic case where successful enzyme-ligand encounters are limited by random diffusion. In this chapter, first the general processivity mechanism is introduced and particular examples of processive enzymes are discussed (Sec. 2.1). Hyaluronidases, glycan degrading processive enzymes, and their main properties, are then described in Sec. 2.2. In Sec. 2.3 we discuss in detail the structure of *Streptococcus Pneumoniae* Hyaluronate Lyase (*Spn*Hyal) along with the catalytic mechanism degrading the glycosaminoglycan substrate Hyaluronan (HA). This enzyme represents a prototypic processive enzyme and all the work presented in this thesis, is carried out using *Spn*Hyal-HA structure complex. Preliminary work carried out using CONCOORD simulations is also sketched briefly.

2.1   Processive Enzymes

Biopolymers — either as polynucleotides (DNA or RNA), polypeptides (proteins), or polysaccharides (cellulose, starch or hyaluronan) — play a central role in biology. For their function and metabolism they often need to be modified. Many enzymes have evolved to synthesise, degrade or modify biopolymers, e.g., DNA- and RNA-synthases, DNA polymerases, the ribosome complex (protein synthesis), proteases (protein degradation), or enzymes acting on glycans such as cellulases (cellulose metabolism), amylases (breakdown of starch) and hyaluronidases. Because of the polymeric/repetitive nature of these biopolymers, many enzymes also act on these polymers processively. Such enzymes are called 'processive' enzymes.

In general, the time required between two successive rounds of catalysis is limited, at least in part, by the frequency of enzyme-substrate encounters. In a non-processive enzymatic mechanism, this frequency depends on diffusion (often random) in three spatial dimensions (or even six, when rotation is also considered). For processive enzymes, however, the substrate remains bound to the enzyme between multiple rounds of catalysis. Thus, instead of a six-dimensional diffusion, the enzyme in this picture slides along the polymeric substrate in a quasi one-dimensional fashion, thereby enhancing the efficiency of the processive enzymes manifold as compared to normal enzymes.

Processive enzymes are of interest not only because of their abundance, but also due to their enormous importance in many human diseases. To give particular examples, recently, major clinical interest has emerged in the blocking of the processive reverse transcriptase involved in HIV infection[173] and enzymes such as protein farnesyltransferase and telomerase that are involved in dominant forms of cancer.[174,175]

Processivity in enzymes was first discovered during the 1960's from studies of the enzyme ribonuclease.[176] Since then, many such proteins have been identified in nature, and techniques have been developed and refined to quantitatively characterise the phenomenon of processivity.[177,178] Processive enzymes can be classified into two main structural groups:[18]

> **Class I. Closed structures:** These enzymes completely enclose the substrate to facilitate the sliding phase between multiple rounds of catalysis. The complete enclosure can again be formed from two kinds of structures:
>
> > **Class I(a), Asymmetric structures:** In some processive enzymes, a single polypeptide chain creates an asymmetric structure that is capable of completely enclosing a polymeric substrate. For example, in Fig. 2.1 (a), the bulk of the enzyme forms a well-defined groove within which the substrate binds. Examples for this type of processive enzyme include exonuclease I from *E. coli* (PDB code: 1FXX), Phage T7 DNA polymerase with processivity factor, and *E. coli* thioredoxin (PDB code: 1T7P). The lytic transglycosylase Slt70 from *E. coli* (Fig. 2.1(a)) processively degrades bacterial-cell wall. It consists of an asymmetric ring, which is created from 22 $\alpha-$helices. Attached to this

ring, via a linker domain, is the catalytic domain, which is structurally homologous to goose egg-white lysozyme. It is thought that the ring allows the protein to encircle the polysaccharide strands that constitute the cell wall, thus confering processivity.

Class I(b), Toroids: These are protein structures that encircle their substrates by forming symmetric, oligomeric toroids. This class includes the sliding clamp proteins that associate with different DNA polymerases to enable processive replication, the eukaryotic protein PCNA, and bacteriophage T4 gp45. Fig. 2.1(b) shows the bacteriophage λ-exonuclease, that processively degrades one strand of double-stranded DNA (dsDNA) from the 5' to the 3' direction. The phage λ uses this exonuclease to facilitate genetic recombination with the *E. coli* chromosome via the double-stranded break repair and single-stranded annealing pathways. λ-exonuclease is a trimeric protein with the three subunits arranged to form a toroid. The channel formed by λ-exonuclease is tapered such that dsDNA may enter one side but only single-stranded DNA (ssDNA) may exit from the other side. Similarly, polynucleotide phosphorylase (PNPase), which plays a critical role in the degradation of mRNA, processively removes nucleotides from the 3' end of single-stranded RNA (ssRNA). The toroidal shape plays a very important role in the processivity of these enzymes.

Class II. Partially enclosing structures: The structures of this class of processive enzymes have a well-developed groove or cleft that partially encloses the substrate, and the enzyme can be described as having the shape of a saddle. Several examples of partially enclosing processive enzymes can be given such as uracil DNA glycosylase (UDG) responsible for processively scanning and removing misincorporated uracil bases from DNA, and AP-endonuclease (APE-1), or the TATA-box binding protein (TBP), required for accurate initiation of transcription in eukaryotes. Similarly, the structures of Taq RNA polymerase and yeast RNA-polymerase-II suggest that a saddle (or claw) shape allows these molecules to remain bound to their substrate, and this "clamp" subunit is critical for processivity of transcription.

Processive DNA helicases, like PcrA, fall into this structural class as they have a saddle-shaped groove to bind ssDNA. Helicases constitute a special case of processive enzymes as they couple the hydrolysis of ATP to the physical movement along their polymer substrate. Helicases use the energy of ATP hydrolysis to open the hydrogen-bonded double helix of DNA, a prerequisite for their movement along the DNA strand. Many cellulases that can catalyse both endo- and exocleavage contain large grooves. These enzymes have loops that fold over their active sites, resulting in an enclosed substrate-binding tunnel, leading to their dual specificity. However, not all proteins with surface groove are processive enzymes, and hence additional information is needed to understand how processivity is accomplished.

Figure 2.1(c) shows hyaluronate lyase from *Streptococcus pneumoniae* (*Spn*Hyal).

The enzyme processively degrades polymeric hyaluronan (HA),[179] yielding only disaccharides as products. The active site of this enzyme resides in a long positively charged groove, which provides a suitable complement to negatively charged HA. In the present work the processivity mechanism in *Spn*Hyal is studied using MD simulations.
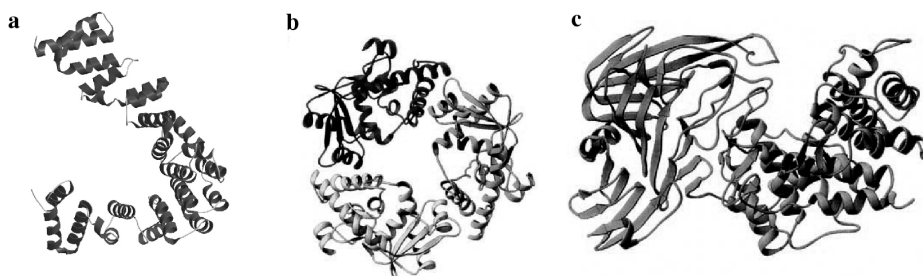
Figure 2.1: Representative structures for each class of processive enzymes. a) Class Ia: Lytic transglycosylase Slt70 (PDB code: 1Q5A); b) Class Ib: λ-exonuclease (PDB code: 1AVQ); and c) Class II: Hyaluronate lyase (PDB code: 1LOH).

Structural studies have provided first hints on the mechanism of processivity for several proteins, but atomic details of particularly the substrate sliding phase between subsequent rounds of catalysis still remain unknown. We try to address this issue with molecular dynamics by elucidating the microscopic mechanism of processivity of a prototypic processive enzyme: hyaluronate lyase (Fig. 2.1(c)), a glycan degrading enzyme of which high-resolution structures from two different species were recently solved — both in the *apo* form of the enzymes, as well as in complex with substrate and product. In the present work, we examine the translocation of the hyaluronan (HA) substrate within the hyaluronate lyase to gain further insight into the mechanism of processivity. Hyaluronate lyases have application in structural studies of connective tissue glycosaminoglycans and have very important physiological and medical roles, which will be discussed in Sec. 2.2.1.

2.2 Hyaluronidases

Hyaluronate lyases (Hyal) fall into the category of partially enclosing processive enzymes. They primarily degrade hyaluronan (HA), the polymeric glycosaminoglycan (GAG). Hyaluronan is a polymer built of disaccharide units, which is present in essentially all higher organisms and also in some microbes (see Chapter 3 for details on HA). The end-product of the degradation is the unsaturated disaccharide unit (for the detailed catalytic mechanism see Sec. 2.3.2)

Gram-positive organisms often secrete Hyals, which are thought to have an important role in pathogenesis, since all the Gram-positive bacteria that produce Hyals are capable of causing infections in animals.[180] Many pathogenic streptococci produce extracellular Hyals, which aid the organism to spread into the host tissues, by processive

degradation of host HA (which forms the connective tissue structure or practically the physical defensive wall of the host) and other glycosaminoglycans.[181] One such gram-positive bacterial organism is *Streptococcus pneumoniae*, a human pathogen that can cause life-threatening diseases such as pneumonia, bacteraemia, and meningitis. Hyals are displayed on the surface of gram-positive organisms and are thus involved in direct interactions with host tissues (see Fig.2.2). With the degradation of hyaluronan Hyals allow microbial access to, or migration between host tissue. Hyals are therefore thought to contribute significantly to pathogenesis and are considered to be important pneumococcal virulence factor to be studied as an alternative for a pneumococcal vaccine. In the later sections we discuss these enzymes in detail.
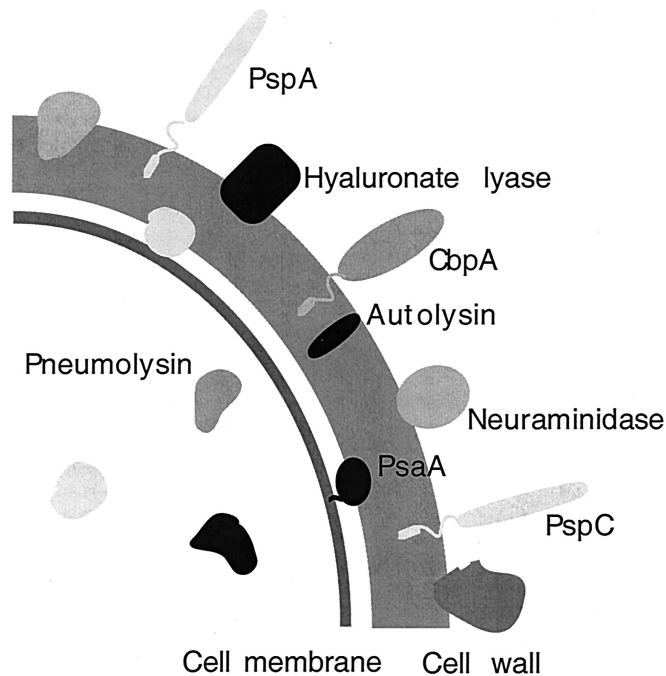
Figure 2.2: Schematic diagram of the virulence factors of *Strep. pneumoniae*. Hyals (represented by black rectangle) are one of the major surface proteins.).

2.2.1 History and physiological importance

Studies on extracts of mammalian testes and other tissues in 1928, showed that there is a 'spreading factor' facilitating diffusion of antiviral vaccines, dyes and toxins injected subcutaneously.[182] In 1937, K. Meyer, who also isolated the structure of hyaluronan (see Chapter 3), identified an HA-degrading enzyme in bacteria[183] and soon the mammalian 'spreading factor' was shown to be an HA-degrading enzyme.[184] The term "hyaluronidase" was introduced by K. Meyer in 1940 to denote enzymes that degrade hyaluronan. Hyaluronidases are widely distributed in nature and many hyaluronidase-

like enzymes* have been detected and/or isolated from a large number of tissues and organisms, e.g. liver, kidney, placenta, uterus, testes, spleen etc.; from the venom of snakes, lizards, bees, scorpions, fish, wasps, spiders etc.; as well as from some bacteria, fungi and invertebrate animals.[187–189] Because they are difficult to purify and isolate, hyaluronidases - although very important - were not studied in detail and had been disregarded enzymes until recently improved techniques were available.[190–192]
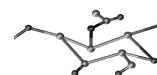
Hyaluronidases are of great importance in many medical fields† like orthopaedia, surgery, ophthalmology, internal medicine, dermatology, and gynaecology.[188,193] Their ability to cleave hyaluronan in tissues results in increased membrane permeability, a reduced viscosity, and a facilitated diffusion of injected fluids. These phenomena, referred to as 'spreading effect' of hyaluronidases, are used to accelerate and increase absorption of injected drugs, e.g. antibiotics, to promote resorption of excess fluids, to improve the effectiveness of local anaesthesia, and to diminish pain due to subcutaneous or intramuscular injection of fluids. Sperm hyaluronidase is involved as a key player in successful fertilisation in most mammalians, including humans.[194] Hyaluronidases are also shown to have positive effects in cancer treatment and chemotherapy.[195,196] With respect to all these applications, the number of pharmacological studies of hyaluronidases to further support the physiological and pathophysiological role of the enzyme and its substrate hyaluronan have developed tremendously in the past decade.[197–202] However, the molecular mechanism behind the hyaluronan degradation by hyaluronidases is still not completely understood. In the present work, we take the first steps towards solving this problem.

2.2.2 Classification of Hyaluronidases and Streptococcus pnumoniae Hyaluronate Lyase

In 1971, K. Meyer introduced a classification scheme for hyaluronidases based on biochemical analysis of the enzymes and their reaction products, which is still remarkably accurate as seen with the advent of genetic data.[203] K. Meyer identified three principal types of hyaluronidases:

1. Hyaluronidases (endo-beta-N-acetylhexosaminidases) degrading hyaluronan by cleavage of the $\beta-1,4-$glycosidic bond leading to tetrasaccharides and hexasaccharides as the major end-products. These are mainly mammalian-type hyaluronidases (EC 3.2.1.35) and the best known enzymes are the testicular, the lysosomal and the bee venom hyaluronidases. These enzymes have both hydrolytic and transglycosidase activities,[204] and can degrade HA and chondroitin sulfates (ChS), specifically C4-S and C6-S, as well as, to a small extent, dermatan sulfate (DS).

---

*These hyaluronidases differ in their molecular weight, substrate specificity and pH optima[185,186]

†For a good reference on science and applications of hyaluronan and hyaluronidases, see Glycoforum (2003) website: http://www.glycoforum.gr.jp/science/hyaluronan/hyaluronanE.html

2. Hyaluronidases (EC 3.2.1.36) from leeches, other parasites, and crustaceans are hyaluronate 3-glycanohydrolase or endo-beta-glucuronidases yielding tetrasaccharide and hexasaccharide as end-products. They degrade HA by hydrolysis of 1,3-linkages between $\beta-$D-glucuronate and N-acetyl-D-glucosamine residues.

3. Bacterial hyaluronidases (EC 4.2.2.1 or EC 4.2.99.1) degrade HA and to various extents, CS and DS. They are endo-beta-Nacetylhexosaminidases that operate by a $\beta-$elimination reaction yielding primarily unsaturated disaccharide end-products. Bacterial hyaluronidases are actually hyaluronate lyases[179,190] isolated from various microorganisms e.g. strains of *Clostridium*, *Micrococcus*, *Streptococcus* or *Streptomyces*,[180,205] and differ in substrate specificity[*].

Alternatively, based on the molecular genetic analysis the hyaluronidases can be divided in two main classes according to sequence homology:

1. **Hyaluronidases from eukaryotes**: This class contains eukaryotic hyaluronoglucosaminidases (EC 3.2.1.35) like mammalian hyaluronidases[†], bovine testicular hyaluronidases (BTH), or bee venom hyaluronidases (BVH).

2. **Hyaluronidases from prokaryotes**: Here a variety of hyaluronidases has been decoded (EC 4.2.2.1)[‡]. The best characterised among these are *Streptococcus pneumoniae* (*Spn*) and *Streptococcus agalactiae* (*Sag*) hyaluronan lyases.[179,209,210] It is this *Spn*Hyal that we study here in the present work as a prototypic case for processive enzymes.

2.3 *Streptococcus Pneumoniae* Hyaluronidase (*Spn*Hyal)

Both, *Streptococcus pneumoniae* (*Spn*) and *Streptococcus agalactiae* (*Sag*) hyaluronan lyases, are among the best studied hyaluronidases.[179,209,211,212] Both cleave the $\beta-1,4-$glycosidic bond of hyaluronan between D-glucuronic acid and N-acetyl-D-glucosamine in a $\beta-$elimination reaction resulting in the unsaturated disaccharide 2-acetamido-2-deoxy-3-O-($\beta-$D-gluco-4-enepyranosyluronic acid)-D-glucose.[179,209,211,212] With the recently resolved very high resolution x-ray structure of *Spn*Hyal[22] at about 1.7Å as well as further biochemical and computational analysis[21,201,209–211,213–221] this enzyme is structurally well-characterised.

---

[*]Some hyaluronidases do have absolute specificity for hyaluronan such as those from the bacteria *Pneumococcus* and the mold *Streptomyces*.

[†]With the tremendous progress in molecular genetics in the last years and the knowledge of the complete "Human Genome"; six-hyaluronidase-like sequences were identified with about 40% sequence identity: Hyal-1, Hyal-2, Hyal-3, Hyal-4, HYALP1 (a pseudogene) and PH20 (SPAM 1).[206,207]

[‡]See e.g. Carbohydrate-Active Enzymes server[208] at URL: http://afmb.cnrs-mrs.fr/CAZY/, 1999.

2.3.1   Choice of *Spn*Hyal for processivity study

Streptococcus pneumoniae hyaluronate lyase (*Spn*Hyal) is expressed in many strains of *Streptococcus pneumoniae* that are involved in several diseases such as pneumonia, bacteraemia, meningitis, otitis and sinusitis.[14, 181, 222]   In the present study, we have chosen *Spn*Hyal as a prototype for processive enzymes for the following reasons:

- High-resolution structures are available for two streptococcal species: *Strep. pneumoniae* and *Strep. agalactiae* for both the native enzyme (*apo* form) and in complex with substrate molecules (*holo* form).[20, 212]

- Its processive mechanism is presumably prototypic for many other processive enzymes in the category of partially-enclosing structures (possibly including nucleases, DNA polymerases and cellulases). However, the processivity mechanism of *Spn*Hyal is non-trivial, and also presents a great challenge. This is because, while an appropriately shaped active site of an enzyme appears desirable for processivity, it is not sufficient and not all the proteins with a substrate-binding cleft present are processive.

- *Spn*Hyal is biochemically well-controlled, and an expression-system is available.[214] The studies presented in this work are carried out in collaboration with the Jedrzejas group at Children's Hospital and Research Institute, USA; which provides both the expertise and infrastructure for a functional and structural characterisation of the simulation results, such as the predicted effect of Hyal point mutants.

- A detailed molecular mechanism for the catalytic process of the enzyme is available.[20, 211]

- There is increasing clinical interest because of the involvement of Hyal in bacterial pathogenesis (see Sec. 2.2.1).

- Previous simulation studies with *Spn*Hyal show that a molecular dynamics study of processive mechanism is feasible[22, 23] (for details see Sec. 2.3.3).

2.3.2   Structure of *Spn*Hyal and its catalytic mechanism

Structure of *Spn Hyaluronate Lyase:*

The structure of *Streptococcus pneumoniae* hyaluronate lyase (*Spn*Hyal) allowed the first insight into bacterial degradation of HA.[211] The full protein consists of four domains: an N-terminal HA binding domain, a small spacer $\beta-$sheet domain, an $\alpha-$domain, and a C-terminal $\beta-$sheet domain, which are connected through peptide linkers. In the crystal structure (PDB code: 1LOH) the first two domains were auto-degraded from the N-terminal region and the structure consists of only two structural domains of approximately equal size connected by a short peptide linker (11 residues). The

(a)

(b)

(c)
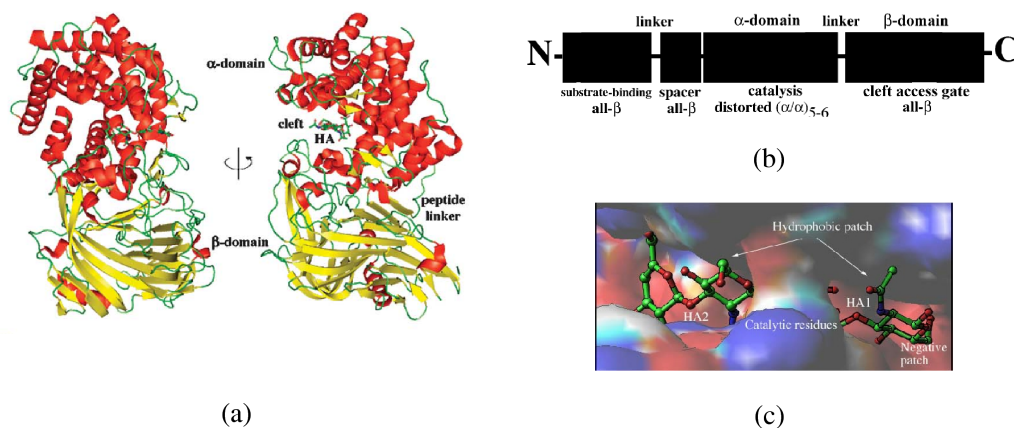
Figure 2.3: Structural details of *Spn*Hyal. a) Two views of the crystal structure of *Spn*Hyal (1LOH.pdb) in cartoon representation. The $\alpha-$helical $\alpha-$domain is in red, while the $\beta-$sheet $\beta-$domain can be seen in yellow colour. The small linker peptide can be seen in the right picture. There HA-binding cleft is clearly seen in the $\alpha-$domain with HA in stick representation. (b) The schematic representation of the domain blocks of the full *Spn*Hyal. In the crystal structure the first two domains at the N-terminus are auto-degraded and the stable structure of the last two domains at the C-terminus with the linker is obtained. (c) Electrostatic potential distribution in the catalytic cleft of the *Spn*Hyal. The majority of the cleft is highly positively charged (shown in blue), whereas the product-releasing end of the cleft (HA1 position) is negatively charged (shown in red). The positions of the hydrophobic (aromatic) patch and catalytic residues are also labeled.

N-terminal $\alpha-$helical $\alpha-$domain consists of 13 $\alpha-$helices (first 361 residues), and the C-terminal $\beta-$domain consists of 24 β strands packed into five antiparallel β sheets (following 347 residues) (See Fig. 2.3). The binding cleft for the HA substrate is located in a broad gorge between these two domains, and is about 30 Å in length and 10 Å in width, thus leaving enough room for three disaccharide units or a hexasaccharide unit (see Fig. 2.3). Most of the cleft belongs to the $\alpha-$domain. Substrate modelling, later confirmed by the x-ray structure in complex with the disaccharide product[20] and tetra- and hexasaccharide units of hyaluronan[22] showed that essentially all the interactions between the protein and the substrate are provided by the $\alpha-$domain.

Basically three types of interactions between *Spn*Hyal and HA substrate were found:

(a) Overall, the binding cleft is dominated by positively charged residues (positive patch), enhancing the affinity to the negatively charged substrate.

(b) Aromatic residues form a hydrophobic patch that matches the sugar rings of the substrate and anchors the substrate precisely into the binding cleft, thereby presumably enhancing substrate affinity and thus selectivity.

(c) A negative patch is located near the binding site of the first disaccharide unit (HA1), which presumably facilitates the disaccharide product release.

PAD mechanism for catalysis:

Based on the studies of the complex structures of *Spn*Hyal with both substrate and product, together with site directed mutagenesis,[20, 22, 211] a new proton-acceptance-donation (PAD) mechanism for HA degradation has been proposed recently*. This mechanism is depicted in Fig. 2.4 both schematically and in detail. The enzyme initially makes a random exolytic bite in the polymer which is then followed by endolytic degradation, one disaccharide unit of HA at a time, until the chain is fully degraded. The steps involved can be summarised as follows:

1. **Binding of Hyal to the polysaccharide HA substrate.** The positive charges in the majority of the cleft guide the HA chain into the cleft. The aromatic patch in the cleft then anchors and positions the cleavage sites on the substrate.

2. **Glycosidic bond cleavage.** The nucleophilic NE2 atom of the His399 residue, assisted by Asn349, accomplishes the removal of the more acidic, axially positioned C5 proton from the glucuronate residue. This results in the formation of an unsaturated C4-C5 bond on glucuronic acid. Tyr408 delivers a proton to the oxygen of the O-C4 glycosidic bond resulting in breaking of this bond. This completes the $\beta-$elimination process of substrate degradation.

3. **Disaccharide product release.** The final degradation product is thus the unsaturated and negatively charged disaccharide unit of hyaluronan (Fig. 2.4(b)). The negative patch in the cleft provides the electrostatic force to release this product making way for the new educt in the next catalytic round.

4. **Proton exchange.** His399 loses its extra proton and Tyr408 obtains a proton by interacting with water molecules in the microenvironment and thus regenerates their original configuration for the next round of catalysis.

5. **Processivity.** The remaining hyaluronan substrate then slides by one disaccharide unit and the process repeats itself until it is completely exhausted.

2.3.3 Previous studies on the processive mechanism of *Spn*Hyals

CONCOORD[223] simulations of *Spn*Hyal both, in the absence and presence of substrate, have provided first clues on the mechanism of processivity in Hyal.[22, 23] Essential dynamics (ED) analyses of the simulated trajectories showed that the collective fluctuations in *Spn*Hyal are dominated by mainly three domain motions. The first type of domain motion (eigenmode 2 in Fig. 2.5) describes an opening and closure of the two domains with respect to each other. This seems to be directly connected to substrate

---

*Previously all the bacterial Hyals have been thought to cleave HA using an elimination mechanism, and hence they were termed hyaluronan lyases, to distinguish them from mammalian hyaluronidases, which are hydrolases, cleaving HA by a hydrolysis process.
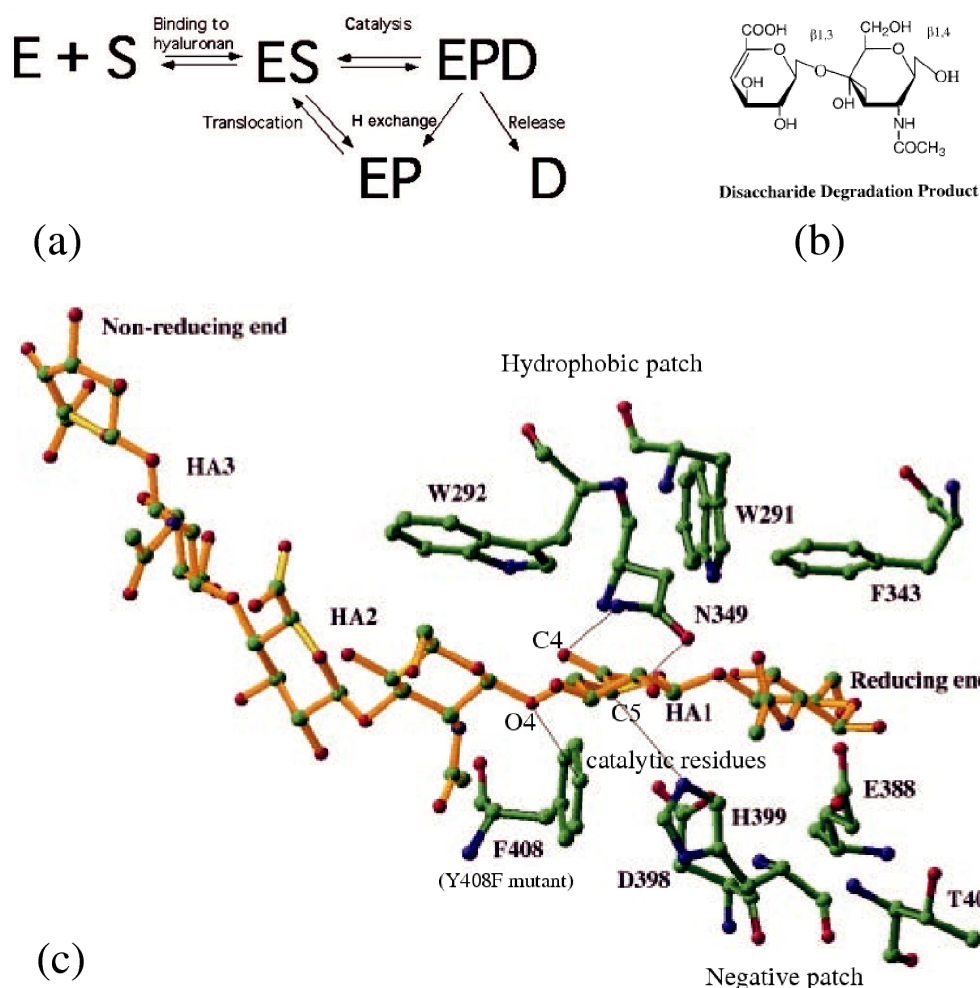
Figure 2.4: (a) A schematic representation of the catalytic mechanism (proton acceptance and donation mechanism or 'PAD') and the processivity in the *Spn*Hyal-HA complex. (b) The unsaturated disaccharide product of the catalytic action between *Spn*Hyal and polymeric HA-substrate. (c) A detailed *Spn*Hyal-HA interface. The hexasaccharide segment of HA (labeled HA1 to HA3) is coloured in orange. The green residues form the catalytic, aromatic and negative patches in the cleft. During the catalytic mechanism an unsaturated disaccharide product is formed at the reducing end.

affinity because for the substrate inside the cleft this opening/closing of the domains markedly changes the contacts with the protein. Although this mode is important for the rounds between catalysis and also during the catalysis, it does not give any insight into how the substrate can actually slide along the cleft between two successive rounds of catalysis. Here two other modes of protein motion have been proposed to play a role. These are a twisting of the two domains (eigenmode 1 in Fig. 2.5) and an opening/closing motion of the entry/exit sites of the cleft (eigenmode 3 in Fig. 2.5). The twisting mode
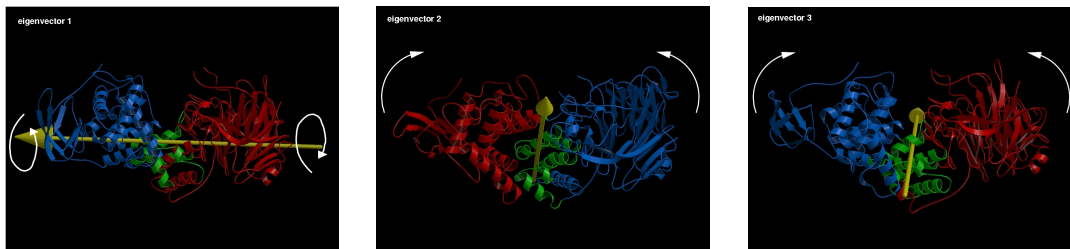
Figure 2.5: The first three eigenmodes (eigenvectors) of the domain motions from the ED analyses of CONCOORD simulations. These represent the three largest domain motions occurring in the protein, namely opening/closing of the cleft (center), twisting of the $\alpha-$domain with respect to the $\beta-$domain (left) and the opening/closing of the access to the cleft (right).

is hypothesized to play a major role during the sliding phase of the substrate, because an analysis of protein-residue contacts for the two extreme structures along this mode showed a shift of about 11Å, which closely matches with the length of a disaccharide unit of HA (about 10Å). This mode therefore was predicted to be involved in the processivity mechanism of this enzyme.

In the present work, therefore, we investigate these domain motions with explicit MD simulations and analyse their influence on the protein-substrate motion in detail. This study may represent an important step towards understanding the processivity mechanism structurally and energetically of *Spn*Hyal in particular, and of its class of partially-enclosing processive enzymes in general.

Chapter 3

Hyaluronan: Structure and dynamics

Summary

Hyaluronan (HA) is a very simple biopolymer containing only two kinds of sugar units. Hundreds or thousands of these sugar units are linked by alternate $\beta-1,4$ and $\beta-1,3$ glycosidic bonds. The present chapter starts with a brief introduction of HA (Sec. 3.1). The OPLS-AA force field has proved to be suitable for the description of proteins in a solvent environment and therefore we chose this force field for all the simulations of *Spn*Hyal in the present work. Since HA is a carbohydrate, we first investigate, in Sec. 3.2, whether this force field is also valid for this molecule. It turns out that some modifications of OPLS-AA force field were needed to describe HA adequately. Thus, we chose this modified set of the force field parameters for HA, throughout the present work.

3.1 Hyaluronan (HA): Introduction

In 1934, K. Meyer and J. Palmer showed that the substance isolated from the vitreous of bovine eye contained an uronic acid and an aminosugar, but no sulfoesters. For this novel glycosaminoglycan, they proposed the name "hyaluronic acid"[224] (from hyaloid (vitreous) + uronic acid). Today this macromolecule is most frequently referred to as 'Hyaluronan' (HA) and is one of the most versatile macromolecules in nature. Hyaluronan is now known to be present in essentially all higher organisms and also in some microbes. In recent years, because of its viscoelastic properties, hyaluronic acid has found a great many medical applications and has become one of the most important biopolymers in this regard.

3.1.1 Chemical structure and polymer properties

It took additional 20 years before the chemical structure of the basic disaccharide motif forming hyaluronic acid was finally determined by Meyer[225] and coworkers. Chemically, hyaluronic acid is a very simple linear polymer consisting of repeating disaccharide units linked through alternating $\beta - 1, 4$ and $\beta - 1, 3$ glycosidic bonds. The uronic acid and aminosugar in the disaccharide are D-glucuronic acid and D-N-acetylglucosamine respectively as shown in Fig. 3.1. Both sugars in the disaccharide unit are spatially similar to glucose, which in the beta configuration allows all of its bulky groups (the hydroxyls, the carboxylate moiety and the anomeric carbon on the adjacent sugar) to be in sterically favourable equatorial positions while the small hydrogen atoms occupy the sterically less favourable axial positions. Thus, the structure of the disaccharide shown in Fig. 3.1 is energetically very stable.

The number of repeat disaccharides in an HA molecule varies greatly and can reach 10, 000 or more resulting in a very high molecular weight upto ~ 4 million daltons (~ 400 daltons per disaccharide). With the average length of a disaccharide of ~ 1 nm, HA polymers of 10, 000 repeats may extend to 10 μm if stretched from end to end.

D-Glucuronic Acid   N-Acetyl Glucosamine
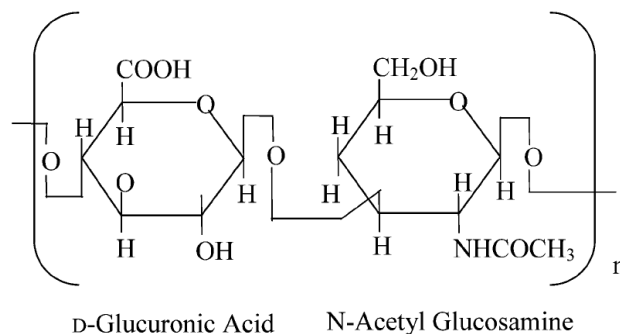
Figure 3.1: The repeat disaccharide of hyaluronic acid, –D-glucuronic acid–β-1,3-N-acetylglucosamine-β-1,4–. The number of repeat units can be n = 20–20000.

Hyaluronan is a member of the glycosaminoglycan (GAG) family, which contains mostly linear polymers of high molecular weight. Other important members of this family are chondroitin/chondroitin sulphates (Ch/ChS), Keratan- and dermatan sulphates, heparin and heparin sulphates. Different from all other glycosaminoglycans HA is strictly composed of non-sulphated sugar units as shown in Fig. 3.1, and to date, no naturally occurring chemical variants of this molecule have been found. Also, in contrast to other GAGs, HA is not covalently bound to a protein core. In 1986, Balasz *et. al.*[226] suggested the term hyaluronan reflecting the fact that it exists *in vivo* as a polyanion due to the mostly charged carboxyl groups of the glucuronic acid residues (pKa = 3–4, depending on the ionic conditions[227]) and not in the protonated acid form. This term has now substituted the older terms hyaluronic acid and hyaluronate.

3.1.2 Solution structure and aggregation properties

There has been long debate about the solution conformation of hyaluronan. Historically, it was assumed for a long time that hyaluronan chains were expanded random coils in physiological solutions. NMR studies performed by Scott *et.al.*[228] for the first time suggested an ordered structure of HA in aqueous solution. Because of 180° rotations between alternating disaccharide units, this conformation model is characterised by a gently undulating, tape-like, two-fold helix, and is stabilised via internal hydrogen bonds and interactions with solvent[229] (Fig. 3.2). The striking feature of this secondary structure of hyaluronan is an extensive hydrophobic patch of about 8 CH-groups from 3 carbohydrate units. Thus, HA includes properties of highly hydrophilic material simultaneously with hydrophobic patches, which is characteristic of lipids. A more recent study[230] with residue-specific NMR and computational studies shows however, that instead of strong intra-molecular hydrogen bonds, weak transient hydrogen bonds that are in rapid interchange with solvent molecules stabilise the solution structure of the HA.

Moreover, HA in an aqueous solution shows self-aggregation[229] into strands of a honeycomb meshwork, where the thickness of the strands increases with HA concentration (see Fig. 3.3). According to $^{13}C-$NMR studies,[232] this aggregation is stabilised both by hydrophobic interactions between the hydrophobic patches and by hydrogen bonds between acetamido and carboxylate groups of neighbouring HA chains arranged antiparallel to each other. Since such hydrophobic and hydrophilic "bonds" can be formed on both sides of the HA polymer, higher order aggregates can assemble causing strands of increased thickness in a HA meshwork. Recently, NMR spectroscopy investigations[233] demonstrated that the tertiary structures (aggregation of antiparallel HA chains) are specifically and reversibly disaggregated by mild physicochemical changes (raising temperature or pH). Further, it was highlighted that the supramolecular organisation of HA is on the edge of stability indicating that reversible formation and breakdown of tertiary structures might have a major influence on controlling biological properties.

These aggregation properties lead to high viscosity, and the volume of hydrated-HA is about 1000-fold increased compared to that of non-hydrated HA.[234] In the hydrated state, the diffusion of *e.g.* proteins and electrolytes is considerably facilitated. Generally,
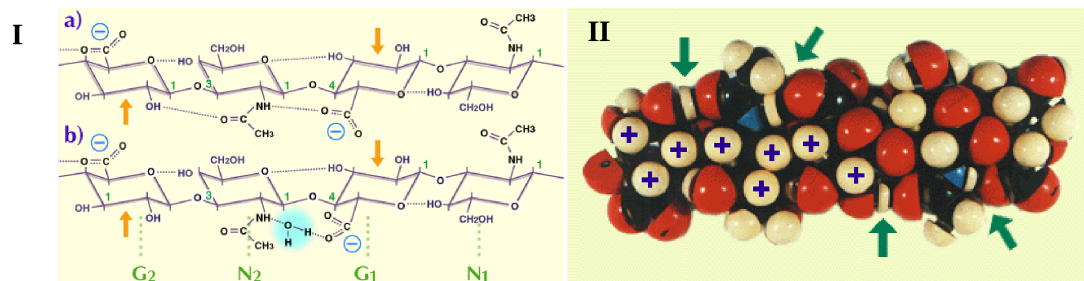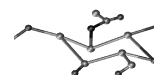
Figure 3.2: I: A tetrasaccharide from a hyaluronan chain, consisting of two disaccharide repeating units showing the preferred configuration in water. G = glucuronate, N = N-acetylglucosamine. The dotted lines indicate hydrogen bonds, of which 5 are potentially present in each hyaluronan tetrasaccharide. (a) aqueous solvent, (b) non-aqueous solvent dimethyl sulfoxide. In non-aqueous solvent the water bridge between N2 acetamido and G1 carboxylate in (b) is replaced by a direct H-bond (a). The G1, N1 disaccharide is rotated 180 degrees about the axis of the chain, compared with the disaccharide G2, N2. The chain is thus a two-fold helix. Arrows indicate the glycol groups, which are resistant to periodate oxidation.
II: Model of a hyaluronan oligosaccharide constructed from space filling Courtauld atoms, based on the structure in I. The H-bonds marked with arrows are between the acetamido and neighboring glucuronate units. The H atoms marked with a cross are part of a hydrophobic patch consisting of 8 CH groups. The hydroxymethyl group can be rotated easily to add its CH2 to this patch. (All figures from Scott, Glycoforum[231])

all molecules can pass through this network, but with different velocities depending on their hydrodynamic volumes. Thus the hyaluronan network acts as a diffusion barrier *in vivo* and may regulate the transport of other substances through the intercellular space.[235] The organisation of HA is supposed to substantially influence the binding of HA to receptors.

3.1.3 Occurance and physiological importance

Found as a major constituent of the extracellular matrices, hyaluronan is present in essentially all higher organisms and also in some microbes. HA is found in the vitreous body of the human eye (0.1–0.4 mg/g wet weight), in synovial joint fluid (3–4 mg/ml), in umbilical cord ($\sim$ 4 mg/ml), in rooster comb (up to 7.5 mg/ml), in the matrix produced by the cumulus cells around the oocyte prior to ovulation ($\sim$ 0.5 mg/ml) or in the pathological matrix that occludes the artery in coronary restenosis. The largest amount of hyaluronan (7–8 g per average adult human, about 50% of the total HA in the body) resides in the skin tissue where it is present in both the dermis ($\sim$ 0.5 mg/g wet tissue) and the epidermis ($\sim$ 0.1 mg/g wet tissue). Furthermore, hyaluronic acid serves as a structural element in the matrix, e.g. in hyaline cartilages (1 mg/g wet weight) where it retains aggrecan molecules in the matrix through specific protein-hyaluronan interactions.[227]

Lower concentrations of HA are found in the matrix of other connective tissues such
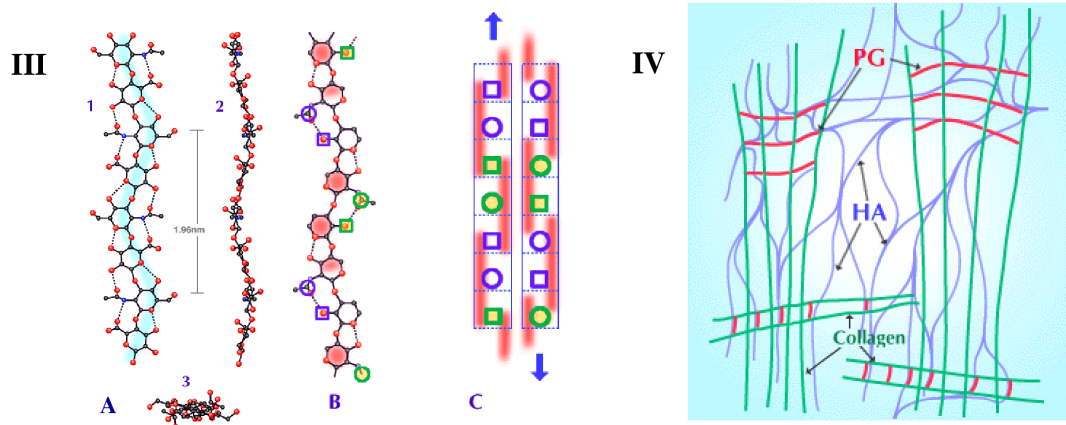
Figure 3.3: I: A – Plan (1) and elevation (2) of computer drawn projections of the tapelike hyaluronan molecule. (3) is the view seen along the two fold helix axis. Note the polymer chain follows a gentle curve in both (1) and (2), in planes at right angles to each other. The N atoms are shown as blue circles. B – Plan projection of a hyaluronan molecule. The red patches are the hydrophobic patches stretching along three sugar units on alternate sides of the polymer chain. Circles represent acetamido and squares represent carboxylate groups. Only the most relevant atoms are shown. C – scheme, seen in side view, of a possible duplex between two molecules of hyaluronan, (similar structures are possible for chondroitin, keratan and dermatan sulfates and mixtures of two or more, including hyaluronan). The two participating molecules are antiparallel to each other. The dotted lines delineate each sugar unit. The red bars are the hydrophobic patches stretching along three sugar units, alternating between front and back of the polymer chain. The blue circles and squares are on the same edge of the molecule and of the duplex (see also B). The curves in each molecule, seen in A and B, closely follow the same course in the antiparallel arrangement so that the two molecules fit well together in the duplex. The hydrophobic patches thus engage closely with each other and the acetamido and carboxylate groups are within H-bonding distances.
II: Scheme of the three-component structure (hyaluronan, collagen fibrils and proteoglycans (PGs)), proposed to be the basis of the stability of vitreous humour. The PGs bridge the collagen fibrils, via specific binding sites on the collagen fibrils, and the collagen–PG structures are separated by a hyaluronan meshwork that interacts with the chondroitin sulfate components of the PG bridges. (All figures from Scott, Glycoforum[231])

as those surrounding smooth muscle cells in the aorta. Hyaluronan as an essential structural element in the matrix plays an important role in tissue architecture by immobilising specific proteins (aggrecan, versican, neurocan, brevican, CD44 etc.) in desired locations within the body. Moreover, hyaluronan is implicated in many biological processes including fertilisation, embryonic development, cell migration and differentiation, wound healing, inflammation, growth and metastasis of tumour cells and whenever rapid tissue turnover and repair are occurring.[234–236] The function of HA may be partly regulated by its chain length, e.g. angiogenesis is presumably induced by small HA oligosaccharides, whereas high molecular weight HA exerts inhibitory effects.[237]

HA interacts with a variety of receptors and hyaluronan binding proteins (hyalad-

herins) on the surface of cells.[235, 238] The great number of hyaladherins known so far can be grouped into (i) the structural hyaluronan-binding proteins of the extracellular matrix, such as link protein and the aggregating proteoglycans, (ii) cell surface hyaluronan receptors and (iii) intracellular hyaluronan binding proteins. The most studied hyaluronan receptor to date is CD44 (lymphocyte homing receptor), which is responsible for a wide variety of cellular functions, e.g. receptor mediated internalisation/degradation of hyaluronan, cell migration and cell proliferation. Several other cell membrane-localised receptors have been identified including the RHAMM (receptor for hyaluronan which mediates motility), ICAM-1 (intercellular adhesion molecule-1), the LEC receptor (Liver Endothelial Cell clearance receptor)[231, 235] and LYVE-1 (Lymphatic endothelial hyaluronan receptor).[239] While most interactions between hyaluronan and hyaladherins are non-covalent the SHAP (Serum-derived Hyaluronan-Associated Protein)-complex represents the only case with covalent hyaluronan-protein crosslinking. The formation of this SHAP-hyaluronan complex plays an important role in the construction and maintenance of certain hyaluronan-rich extracellular matrices.[240]

3.1.4 Medical applications

Probably, in the late 1950s, the first medical application of hyaluronan to humans was a vitreous humour supplement/replacement during eye surgery. Due to hyaluronan's high water-binding capacity and high viscoelasticity, HA is suitable for various medical and pharmaceutical applications. For example, given that HA retains moisture, it is used since a long time in cosmetics.

A rapid increase of hyaluronan levels can occur in many clinical situations, for example during urticaria, the oedema associated with wound healing and inflammation, and the organ enlargement that occurs after transplantation. Furthermore, circulating levels of hyaluronan rapidly increase in situations such as shock, septicaemia and in burn patients.[236, 241, 242]

One of the most successful medical applications of HA is the use of sodium hyaluronate and a covalently cross-linked form of hyaluronan for the treatment of osteoarthritis.[243] It has been reported that sodium hyaluronate suppresses cartilage degeneration, protects the surface of articular cartilage,[244] normalises the properties of synovial fluids[245] and reduces the perception of pain.[246, 247]

3.2 Force field choice

The first question regarding simulations of the HA either free or in complex with the protein was which force field to use. Because OPLS-AA force field[77] has proved to be suitable for the description of proteins in the solvent environment we choose this force field for all the simulations of *Spn*Hyal in the present work. Since HA is a carbohydrate, we first investigate whether the OPLS-AA force field is also valid for this molecule.

3.2.1 Simulation setup

A hexasaccharide segment of the hyaluronan substrate was taken from the *Spn*Hyal-HA complex (pdb code: 1LOH, which was used for all the simulations in the present work). This hexasaccharide segment was then simulated for 100 ns with 6 different forcefield variants (see Table 3.1) and validated against the available experimental data.

No.	force field variant	water model	comments
1	Gromos96	SPC	United Atoms – Gromos96 (kindly provided by P. Hünenberger[248])
2.	AG93	SPC	United Atoms -Amber/Glycam93[68, 249].
3.	OPLS-AA	TIP4P	All Atom – OPLS[69, 70]
4.	OPLS-AA*	TIP4P	All Atom – OPLS with partial charges calculated at DFT level (B3Lyp/6-31G)
5.	OPLS SEI	TIP4P	All Atom – OPLS with molecule specific Scaled Electrostatic Interactions (SEI)[250] (also see Appendix B)
6.	OPLS-SEI*	TIP4P	All Atom – OPLS with both 4 and 5 taken together.

Table 3.1: List of simulations carried out to determine the most suitable force field for hyaluronan simulations. All the simulations were carried out in explicit water solvent with the water model as specified.

All the simulations were carried out with the MD software package GROMACS 3.1.4 and GROMACS 3.2.1.[81] The hexasaccharide segment of the hyaluronan substrate removed from the *Spn*Hyal-HA complex was energy-minimised using the steepest-descent algorithm (Sec. 1.2.6), followed by solvation in a cubic box of approximately 50 x 50 x 50 Å^3 in size containing $12,360$ atoms for the simulations with SPC water model (4091 water molecules), and $15,995$ atoms with TIP4P water model[130] (3961 water molecules). To compensate for the negative charge on the HA hexasaccharide, 3 counterions Na^+ were added.

Each of the fully solvated and neutralised simulation systems was then again energy minimised to obtain the starting configuration for the simulations. The solvent and HA were separately coupled to an external temperature bath[141] of 300 K with a relaxation time of 0.1 ps. In all simulations the system was weakly coupled to a pressure bath of 1 atm with isotropic scaling and a relaxation time constant $\tau_p = 1\ ps$. Bond lengths were constrained to their equilibrium lengths using the LINCS algorithm.[100] This algorithm allows a 2 fs time step for the leap-frog integration scheme. For the Lennard-Jones interactions, a cut off distance of 1 nm was applied. Electrostatic interactions between charge groups at a distance less than 1 nm were calculated explicitly, and the long-range electrostatic interactions were calculated using the Particle-Mesh-Ewald method[147] with a grid spacing of 0.12 nm and a fourth-order spline interpolation (see also Sec. 1.2). A

2 ns MD simulation was carried out with harmonical constraints on the protein heavy atoms with a force constant of $k = 1000\,\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-2}$ to equilibrate water and ions of the system. A subsequent MD simulation of 5 nm length was performed to equilibrate the whole system while the root mean square deviation (RMSD) for the sugar was monitored. Finally, the production run was carried out for 100 ns, storing the coordinates of all the atoms at every picosecond for further analysis. A typical simulation set up is shown in Fig. 1.2 (a).

3.2.2   Validation and choice of OPLS-SEI* force field

All the simulations were validated using two approaches: Principal Component Analysis (PCA) and glycosidic dihedral distribution. For the principal component analysis, first all available x-ray structures of HA were collected. All the simulations were then projected onto the eigenspace of PCA of these experimental data to see how well they overlap with the space spanned by the available experimental structures. Since the PCA on the complete sugar resulted in unrealistic eigenvectors for the simulations like almost complete rotation around the glycosidic bond, PCA was performed on the individual units of the sugar *e.g.* three disaccharide units (D1, D2, and D3) and two tetrasaccharides (T1 and T4). This way it was ensured that the PCA eigenvectors correspond to the true configurational space sampled by both MD simulations and the experimental x-ray structures.

Figure 3.4 shows the main results of these projections. From this figure it can be easily seen that OPLS-AA (red) overestimates the flexibility of this sugar and should not be used. Except pure OPLS-AA, all the other variants are quite in good agreement with the experimental data. The GROMOS (magenta) and the AG93 (black) force field yield quite rigid sugars and some times do not explore the eigenspace sufficiently to sample the experimental data points. The other force field variants show intermediate behaviour if sampling the conformational space for the experimental data is concerned. OPLS with quantum charges modification (green) and OPLS-SEI with the same quantum charges (yellow) show profound effects on the distribution and many times they sample a different space than the others. However, they also agree very well with all the experimental data and sample almost all the experimental structures during 100 ns of simulation time. Although they sample more space than AG93 they do not yield overly flexible sugar as pure OPLS-AA sometimes does. In fact as can be seen from D1-mode1:2, D1-mode2:4, and D3-mode1:4 projections using these force fields explore outlying areas to sample the experimental structures that are not visited by the GROMOS or AG93 force fields. Following these results, we can choose a force field that is compatible to the OPLS-AA force field used for the simulations of the protein (*Spn*Hyal).

Since, the main degrees of freedom in any sugar are its glycosidic bond dihedral transition, the determination of the right force field was also assisted by the analysis of the dihedral distribution of the glycosidic linkages. Figure 3.5 shows the distribution of five dihedral angles in the simulated sugar, and corresponding dihedrals for the experimental structure. For the first and fourth linkage ($\phi 1$ and $\phi 4$, respectively) the OPLS-AA
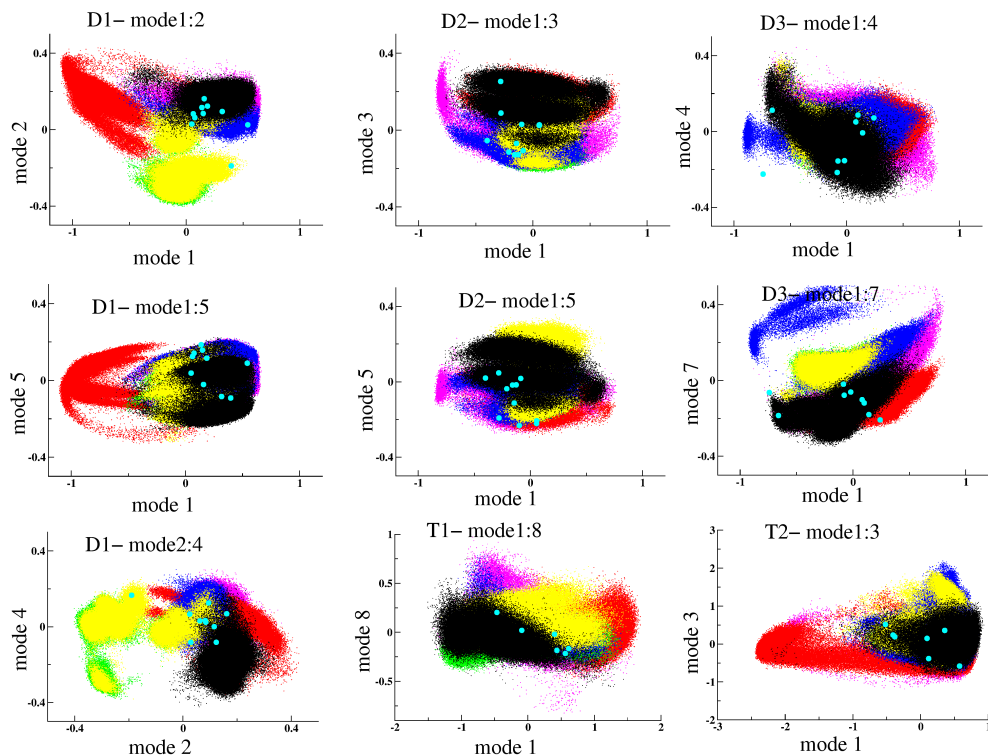
Figure 3.4: 2D projections of all the simulations in Table 3.1 onto PCA-eigenvector sets. D1, D2, and D3 denote that PCA is performed on the respective disaccharide units of the hexasaccharide substrate. Similarly, T1 and T2 denote tetrasaccharide units. Colour code in each plot is as follows — black: United Atoms – Amber/Glycam93, magenta: United Atoms – Gromos96, red: All Atom – OPLS-AA, green: All Atom – OPLS-AA with modified partial charges (OPLS-AA*), blue: All Atom – OPLS with scaled molecule specific electrostatic interactions (OPLS SEI), and yellow: All Atom – OPLS SEI* (OPLS-AA with both SEI and new partial charges), and cyan circles: experimental data.

force field clearly yields a distribution different from the experimental x-ray structures. Only electrostatic interaction modifications (OPLS-SEI) were found not sufficient to alleviate this as is evident from the dihedral distribution in the fifth glycosidic linkage. The remaining force fields result in correct distribution of all the glycosidic linkages. Since the protein in the present work is described by the OPLS-AA force field, the most compatible force field variant was chosen for the HA. This OPLS-SEI* contains both the electrostatic interaction scaling as well as the new partial charges as obtained by quantum mechanical calculations (see Appendix B for the detailed force field parameter set of the HA substrate used in the present work).
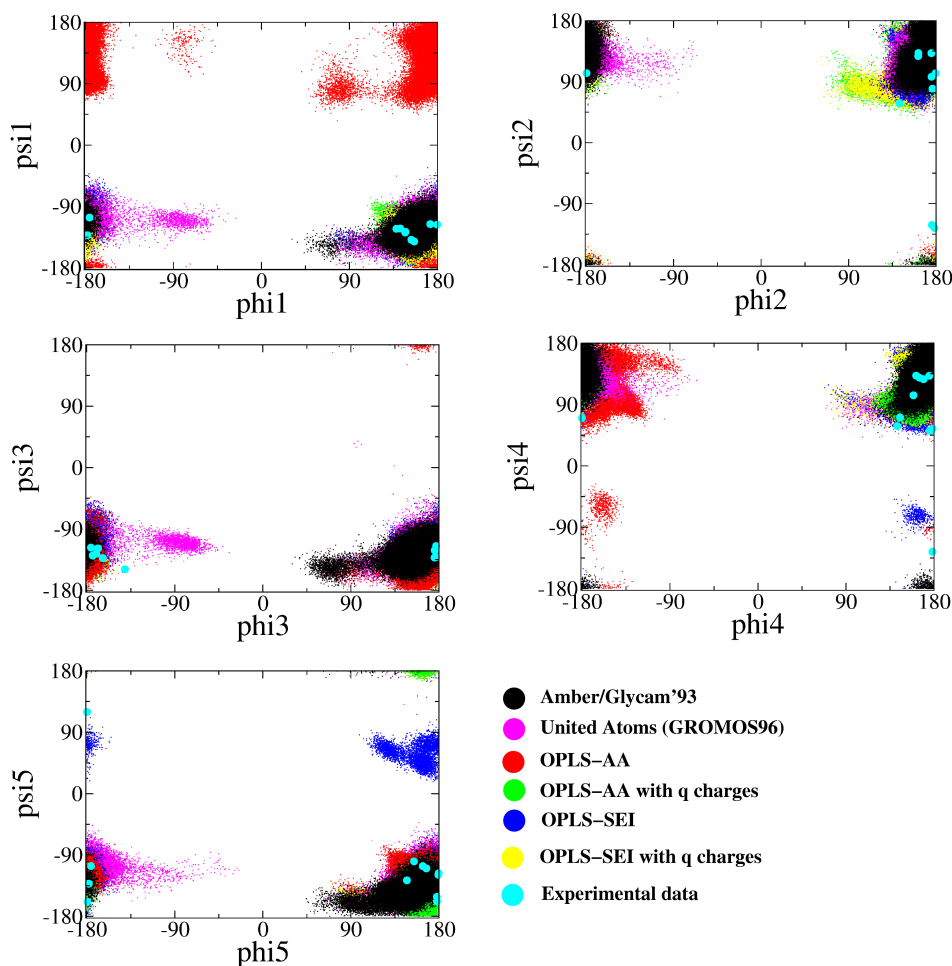
Figure 3.5: Dihedral distribution along the glycosidic linkages of HA6 polymer. All the simulations are validated against the structural data available. The distribution of the dihedrals with pure OPLS-AA force field (red) is clearly does not match with the experimental data for two of the glycosidic linkages. Scaling of the molecule specific interactions (SEI) helps improve the distribution drastically (blue). The chosen force field for the present work (yellow) consists of this scaled parameters as well as modified partial charges of the system as calculated quantum mechanically. (colour code is as explained in Fig. 3.4.)

3.2.3 Flexibility of HA in the presence of the SpnHyal

With the new chosen OPLS-AA force field variant (OPLS-SEI*), the hexasaccharide segment of HA was simulated in the presence of *Spn*Hyal (For details of this simulation setup see Chapter 4). Here, we focus on the difference between the behaviour of the sugar in the presence or absence of the surrounding protein. A particular question to be addressed was whether the behaviour of the $\beta - 1, 4$ linkage with respect to the $\beta - 1, 3$ linkage changes because *Spn*Hyal selectively cleaves $\beta - 1, 4$ glycosidic bond. Thus all
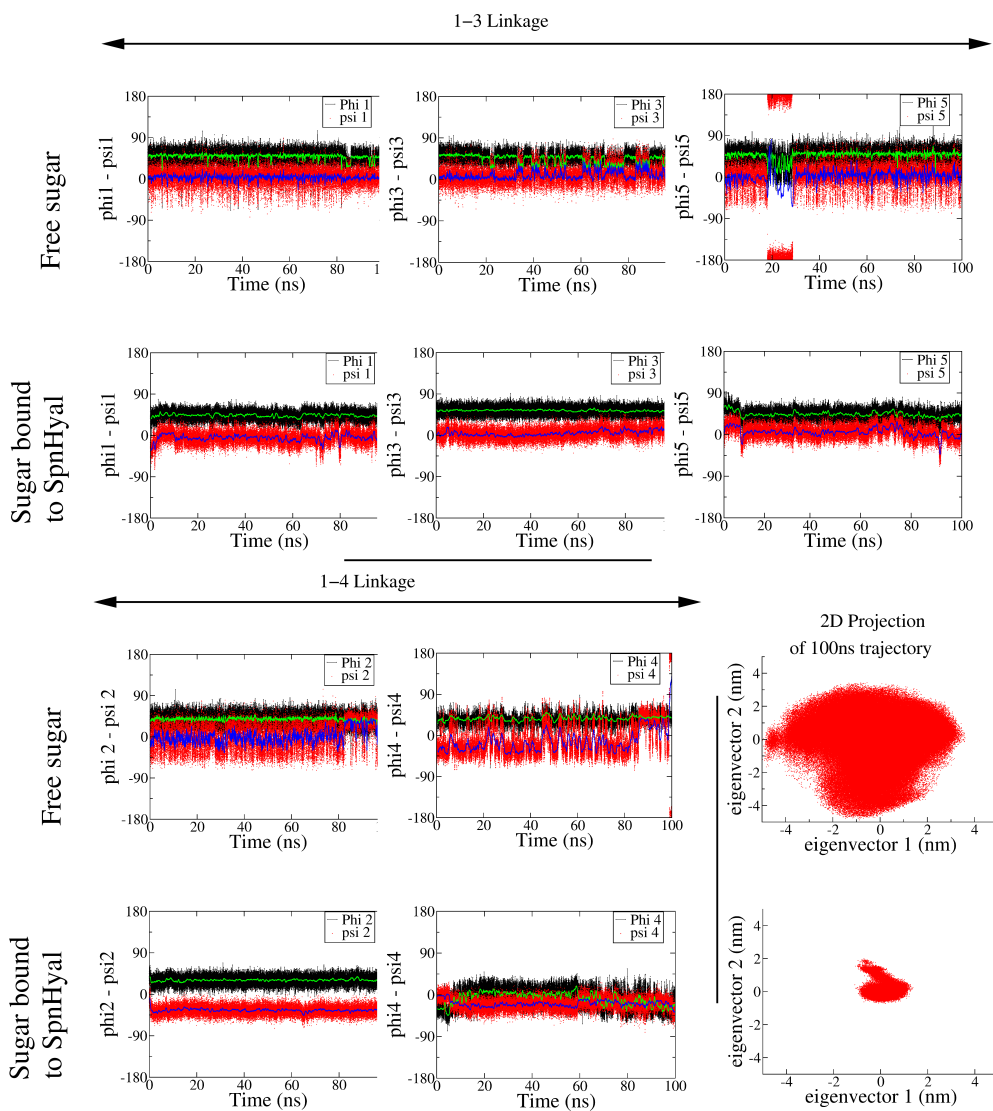
Figure 3.6: Time evolution of dihedral distribution along the glycosidic linkages of HA in free simulation as well as in the presence of *Spn*Hyal. In all the plots, red dots denote the psi distribution along with the blue trace of running average at the interval of 100 ps overlayed, and black lines denote the phi distribution along with the green trace of running average at the interval of 100 ps overlayed. As can be seen the overall flexibility of the sugar along all the glycosidic dihedrals reduces in the presence of the enzyme. This is also depicted in the 2D projection of the first two eigenvectors obtained from the PCA of the sugar.

the glycosidic linkages were observed during the simulations (free sugar and sugar bound to *Spn*Hyal).

Figure 3.6 shows the time evolution of the glycosidic dihedral angles over the simulation time. Generally, HA in free solution shows overall increased flexibility compared to

HA inside the cleft as can be seen in the projection of the first two eigenvectors of the PCA performed on the sugar trajectories of both cases. One major difference between the $\beta - 1, 3$ and $\beta - 1, 4$ is seen in both, free-HA and HA bound to SpnHyal. The $\beta - 1, 4$ glycosidic bond is seen to be more flexible than the $\beta - 1, 3$. It is seen that all linkages of the same type (i.e. either 1-3 or 1-4) show a similar range of motion, irrespective of where they are situated in the oligosaccharide chain. The average ($\phi - \psi$) angles for the 1-3 linkage in free simulation are (46.8^0, 3.9^0) and for the 1-4 linkage are (37.6^0, 9.9^0). In the presence of the protein these averages are (41.3^0, 0.97^0) for 1-3 linkage and (18.6^0, -26.4^0) for 1-4 linkage. This unusual change in 1-4 linkage is mainly due to the phi4psi4 pair as seen from the figure.

3.3 Conclusions

The dynamical dependence of HA on different force field variants was investigated by employing many different long time-scale (100 ns) molecular dynamics simulations. All the simulations are cross-validated against the available experimental data using two approaches – PCA and glycosidic linkage dihedral distribution. Even slight changes in the partial charges show a profound effect on the behaviour of the sugar. An improved version of the OPLS-SEI force field as developed by Kony *et al.*[250] adequately describes the carbohydrates. This force field variant along with modified partial charges as calculated explicitly by a quantum mechanical way has shown better agreement with the available experimental data than normal OPLS-AA force field. AG93 and Gromos96 are also suitable to describe the HA behaviour, however, since we simulate SpnHyal with the OPLS-AA force field, we prefer the variant of this force field for all the simulations of HA in the present work.

HA shows quite dramatic flexibility along its glycosidic linkages. This flexibility was studied for the free and SpnHyal bound HA. For free HA simulation, the $\beta - 1, 4$ glycosidic bond shows a broader distribution in the $\phi - \psi$ space compared to that of $\beta - 1, 3$ glycosidic bond. In the presence of SpnHyal this flexibility is decreased. This is interesting since SpnHyal is known to cleave only $\beta - 1, 4$ glycosidic bond selectively during the catalysis. It may be speculated that because of its inherent flexibility, the $\beta - 1, 4$ glycosidic bond can explore many preferred orientations in the limited space of the catalytic cleft to be cleaved by the protein, which is not possible for the more rigid $\beta - 1, 3$ linkage. More detailed investigations of the dynamics along these $\beta - 1, 4$ and $\beta - 1, 3$ linkages such as studies revealing why these linkages have different flexibility in the free environment may give insights into the structural basis for such specificity.
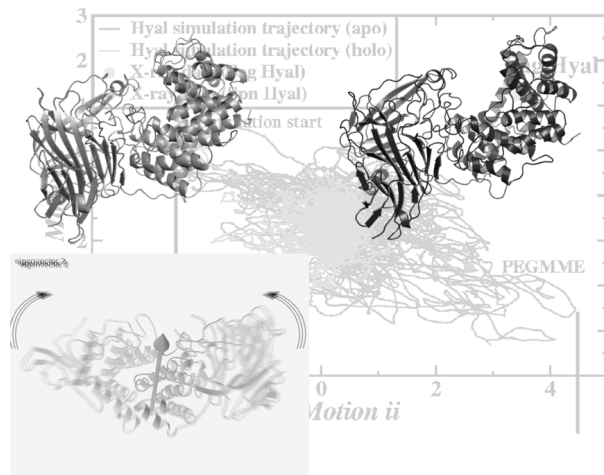
Chapter 4

Flexibility of Hyaluronate Lyase: Part I

Summary

Biomolecules often show large-scale conformational motions that may be functionally essential. To investigate such conformational motions, we employ molecular dynamics simulations of *Streptococcus pneumoniae* Hyaluronate lyase (PDB code: 1LOH). Three principal modes of motions are observed that may be relevant to the processivity mechanism. The main observation is that in the presence of hyaluronan substrate, the protein gets 'locked up' in a closed conformation; while the *apo Spn*Hyal is seen to be more flexible. In this chapter we sketch these observations of dynamical exploration from "closed" to "open" conformation in detail. After a brief introduction in Sec. 4.1, the experimental procedures are sketched in Sec. 4.2, in which two novel structures obtained by the Jedrzejas group are introduced. The simulation set-up is then discussed in Sec. 4.3. The Principal Component Analyses (PCA) performed on these simulations and other analyses are discussed in detail in Sec. 4.4. Finally, the implications of these observations to the overall processivity mechanism of *Spn*Hyal are discussed in Sec. 4.5.

Part of this chapter was published in:

4.1   Conformational motions of proteins

A single, fixed conformation is adequate only to represent a very rigid protein. Such a system would correspond to a "lock-and-key" model in which the protein exists in a single well-defined state. However, the energy landscape of most proteins is frequently described in terms of a folding funnel in which there are many unfavourable states that collapse via multiple routes into possibly several favourable folded states representing minima in the energy landscape. The protein can explore these minima by changing its structure (See Fig. 4.1). The flexibility of a protein is thus an inherent property that is often important for its functionality. The conformational changes can be small chair-boat changes in a sugar ring[251] upto large scale concerted motions in e.g. the ribosome.[252] Well-known examples are molecular motors, like muscle proteins, whose protein dynamics is directly related to the mechanics of muscle contraction. Thus, an investigation of the dynamics of the structural fluctuations and its relation to the reactivity and conformational change of the protein is very important for any attempt to understand the activity of a biomolecule. Against this background, we investigate the conformational motions of *Spn*Hyal and their relations to the HA-binding with MD simulations.
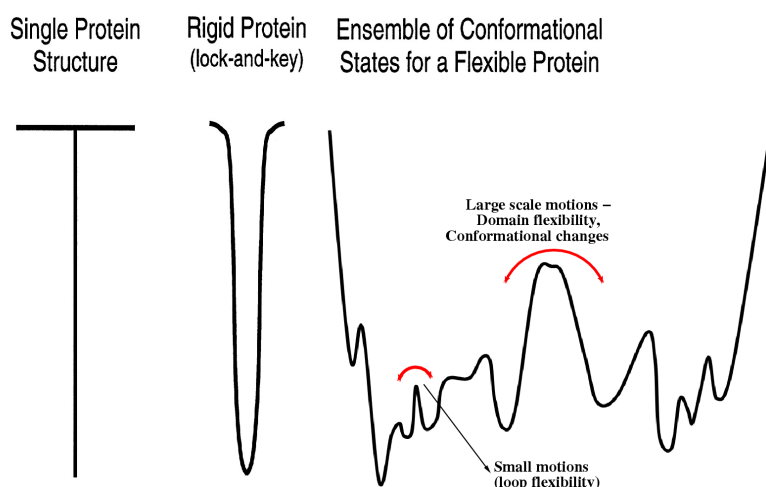
Figure 4.1: A single structure of a protein implies an all-or-nothing folding funnel, perhaps best described with the mathematic singularity shown on the left. The folding funnels on the right and in the center demonstrate the conformational flexibility of a "standard" flexible protein and a "rigid" protein, respectively.

The main motivation behind the present study was to investigate the possibility of the *Spn*Hyal to explore conformations different from the "closed" conformations that were available from x-ray crystallography.[20, 21, 211, 213–215] Apart from these closed conformations of *S. pneumoniae* (*Spn*Hyal), open conformations were available for Hyal from a different species: *S. agalactiae* (*Sag*Hyal), and it was speculated that domain opening was an inherent feature of the Hyal structure. However, it was unknown why

only the *S. agalactiae* Hyal structures were found in an open configuration. To investigate domain opening in *Spn*Hyal, MD simulations that focus at the domain flexibility were carried out and the results are presented in this chapter. The simulation results are compared to two novel x-ray structures of *Spn*Hyal that were obtained simultaneously. The implications of the revealed domain motions for the processivity mechanism are also discussed.

4.2 Crystallisation of new SpnHyal structures

The work presented in this chapter was carried out in close collaboration with the Jedrzejas group, where two new structures of *Spn*Hyal were crystallised and solved. Below, we briefly discuss the results from the experimental investigations carried out in Jedrzejas group.

4.2.1 Crystallisation and diffraction data collection

The wild-type *S. pneumoniae* Hyal, in its fully functional truncated form composed of the catalytic and C-terminal domains only, was produced as reported previously.[213, 214] Two new crystal forms were obtained. These crystals grew within several months and were of different morphology compared to those of the wild-type enzyme reported before.[213] The first set of crystals was obtained employing 30% (w/v) polyethylene glycol monomethyl ether (PEGMME) (pH 6.5), and the second set of crystals was obtained utilising 70% (w/v) saturated malonic acid (pH 6.6). These are henceforth referred to as PEGMME and MALONATE structures, respectively.

The crystals were cryo-protected and frozen using standard methodology. The diffraction data were collected using a synchrotron source, beamline 5.0.1 of the Berkeley Center for Structural Biology, Advanced Light Source, Lawrence Berkeley National Laboratory. The PEGMME crystals yielded x-ray diffraction data at 2.8 Å, and indexed to the $P21$ space group with unit cell dimensions $a = 86.4$ Å, $b = 83.3$ Å, $c = 98.7$ Å, and $\gamma = 98.8°$. The MALONATE crystals yielded a 3.3 Å diffraction data set, space group $C2$, and unit cell dimensions of $a = 117.7$ Å, $b = 101.1$ Å, $c = 85.2$ Å, and $\gamma = 125.1°$. The reasons for the relatively poor diffraction by the new crystal form are unclear. The crystals took a very long time to grow, and conceivably, the structural flexibility of the enzyme is reflected in the crystal quality and the resulting x-ray diffraction data.

4.2.2 Structure solution and refinement

The structures were solved by molecular replacement methods employing the MOL-REP programme and utilising the 1.56 Å crystal structure of the wild-type *S. pneumoniae* Hyal structure (*Spn*Hyal) as a search model. This structure was then divided into smaller parts based on the structural and dynamic domains previously identified.[22, 23] This approach was necessary to successfully achieve a solution of the *Spn*Hyal-PEGMME
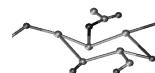
structure. Solutions were further refined using rigid body refinement in CNS software. Computational refinement was alternated with manual rebuilding with O. All data were used throughout with no intensity or sigma-based cut-offs applied. SigmaA-weighted map coefficients and maximum likelihood functions were always used. An R_{free} value, calculated from 5% of reflections set aside at the outset, was used to monitor the progress of refinement. The asymmetric unit of the *Spn*Hyal-PEGMME crystal form contains two Hyal chains, which were constrained to identical conformations with CNS during initial refinement. Later, the constraints were replaced by restraints and, where the density suggested conformational differences between the chains, specific regions were exempted entirely from the restraints. During the final stages of refinement exempted residues were 385-388, 420-433, 435, 444, 492-493, 518, 524 and 527-540. At a late stage, water molecules were placed into 3σ positive peaks in $|F_o - F_c|$ maps if density was also evident in $|2F_o - F_c|$ maps and suitable hydrogen bonding partners were available. Sulfate and malonate ions derived from the crystallisation solutions were modeled into suitably shaped regions of electron density. Temperature factors were refined first on a per-residue basis, and then per-atom for a limited number of iterations guided by the behaviour of the R_{free} value.

4.3   Simulation details

*Spn*Hyal-*apo* and *Spn*Hyal-*holo* were both simulated with explicit MD-simulations. Both simulations started from the closed *Spn*Hyal conformation (PDB code: 1LOH).[22] All the simulations were carried out with the MD software package GROMACS 3.1.4 and GROMACS 3.2.1.[81,82]   For the Hyal-*apo* simulation the HA substrate was removed from the structure manually. The N-terminal and the C-terminal residues (1VAL and 721LEU) were corrected using WHATIF.[253] The protonation states for histidines were corrected manually. The system was then energy-minimised using steepest-descent algorithm (Sec. 1.2.6), followed by solvation in a cubic box of approximately 76 x 102 x 84 $\mathring{A}^3$ in size containing $\sim 80,000$ atoms ($\sim 16,000 - 18,000$ water molecules). The 515 crystallographic water molecules were retained. For the *apo* system 8 counterions Na^+ and for the *holo* system 11 counterions Na^+ were used to neutralise the charge on the system. The water model used was TIP4P[130] and the force fields used were OPLS-AA[69,70] for protein and OPLS-SEI* for the HA hexasaccharide (see Chapter 3).

Each of the simulation systems was then again energy minimised to obtain the starting configuration for the simulations. The solvent, HA, and the protein were separately coupled to an external temperature bath[141] of 300 K with a relaxation time of 0.1 ps. In all simulations the system was weakly coupled to a pressure bath of 1 atm with isotropic scaling and a relaxation time constant $\tau_p = 1$ ps. Bond lengths were constrained to their equilibrium lengths using the LINCS algorithm.[100] This algorithm allows a 2 fs time step for the leap-frog integration scheme. For the Lennard-Jones interactions, a cut off distance of 1 nm was applied. Electrostatic interactions between charge groups at a distance less than 1 nm were calculated explicitly, and the long-range electrostatic

interactions were calculated using the Particle-Mesh-Ewald method[147] with a grid spacing of 0.12 nm and a fourth-order spline interpolation (see also Sec. 1.2). A 2 ns MD simulation was carried out with harmonical constraints on the protein heavy atoms with a force constant of $k = 1000 \, \text{kJ} \, \text{mol}^{-1} \, \text{nm}^{-2}$ to equilibrate water and ions of the system. A subsequent MD simulation of 5 ns length was performed to equilibrate the whole system while the protein backbone RMSD was monitored. Finally, the production run was carried out for 50 ns, storing the coordinates of all the atoms at every picosecond for further analysis. A typical simulation set up is shown in Fig. 1.2 (b). Table 4.1 below shows the details of the two simulations carried out.

Simulation	details	time	System size
*Spn*Hyal-*apo*	1loh structure without HA substrate	50 ns	77 278 atoms
*Spn*Hyal-*holo*	1loh structure	50 ns	83 542 atoms

Table 4.1: Details of two simulation systems. Both the simulations start from the same crystal structure (pdb code: 1LOH). For *Spn*Hyal-*apo* simulation the sugar was manually removed from the crystal structure.

A principal component analysis (PCA) was used to analyse the concerted fluctuations of domains from the MD trajectories (for details on the PCA method see Sec. 1.3.1). To compare the collective motions of simulations of *Spn*Hyal-*apo* and *Spn*Hyal-*holo*, the PCA was first performed on available crystal data including the two new conformations (see Table 4.2). The two simulations were then projected onto the resulting set of eigenvectors for direct comparison of the extent of flexibility.

Abbreviation	Bacterium	Ligand	Resolution (Å)	PDB code
*Spn*Hyal[211]	*S. pneumoniae*	None	1.56	1egu
*Spn*Hyal-*holo*[20, 22]		hyaluronan-		
		disaccharide/	1.70	1c82
		tetrasaccharide/	1.53	1lxk
		hexasaccharide	1.70	1loh
*Spn*Hyal-PEGMME[219]		None	2.80	2brw
*Spn*Hyal-MALONATE[219]		None	3.30	2brv
*Spn*Hyal-vitamin C[21]		Vitamin C	2.00	1f9g
*Sag*Hyal[212]	*S. agalactiae*	None	2.10	1f1s
*Sag*Hyal-*holo*[23]		hyaluronan-		
		hexasaccharide	2.20	1lxm

Table 4.2: Crystal structures referred to in this work and their abbreviations. Starting structure for the simulations is *Spn*Hyal-*holo*. *Spn*Hyal-PEGMME and *Spn*Hyal-MALONATE refer to the newly solved structures while *Sag*Hyal refers to structures from different bacterial species.

4.4 Results

The main feature of the catalytic α-domain is the presence of a large, long cleft where the substrates bind and undergo catalysis. This domain exhibits high flexibility that presumably facilitates the mechanism of action of substrate degradation. From the molecular dynamics simulations this flexibility was analysed in detail and its presumed role in the processivity mechanism was investigated. The MD simulations were compared to each other and to the x-ray structural data available along with the two novel x-ray structures resolved by the Jedrzejas group. Below, we discuss the results from both, the experiments and the simulation studies, in detail.

4.4.1 Description of the new pneumococcal Hyal structures

The individual domain structures of Hyal in the two new conformations, *Spn*Hyal-PEGMME and *Spn*Hyal-MALONATE, are similar to those observed in earlier crystal structures[20,22,211] (Fig. 2.3). The nomenclature PEGMME and MALONATE indicates the main crystallisation agent used to obtain diffracting crystals of *Spn*Hyal in the two new conformations. The crystallised Hyal molecule is composed of two domains, the catalytic domain (α-domain) having an α5/α5 barrel structure and the β-domain comprised mainly of an antiparallel, three-layer β-sandwich). As in earlier crystal structures the catalytic cleft transverses the α-domain. The active site of the enzyme is located within the barrel domain on one side of the cleft area.[21,22,210] Both new structural conformations of the pneumococcal Hyal enzyme display a variety of structural differences from previously observed Hyal crystal structures. These were analysed using intra-domain root mean square deviation (RMSD) calculations and Dyndom to describe domain motions.

The shape of the cleft varies significantly among all structures. In the closed conformation of *Spn*Hyal-*holo* it is even covered by a bridge over the catalytic residues forcing the substrate to thread through the cleft for catalysis (Fig. 4.2 upper left). In *Spn*Hyal-PEGMME the cleft is widely open, in a conformation that would facilitate processivity by allowing easy substrate translocation towards the reducing end.[211,218] The remaining structures appear to illustrate intermediate cleft opening/closing stages. In the *Spn*Hyal-MALONATE structure the cleft is nearly closed and in the *Sag*Hyal structure it is nearly open.

4.4.2 Bacterial Hyals are flexible

The Hyal enzyme exhibits a high flexibility that facilitates the mechanism of action of this enzyme. A previous flexibility analysis revealed three major types of motion of the enzyme (Fig. 2.5):[22,23] (i) a rotation/twisting motion of the whole α-domain relative to the top half of the β-domain (Fig. 2.5 left); (ii) an opening/closing domain motion influencing the width of substrate-binding cleft (Fig. 2.5 center); and (iii) the motion of
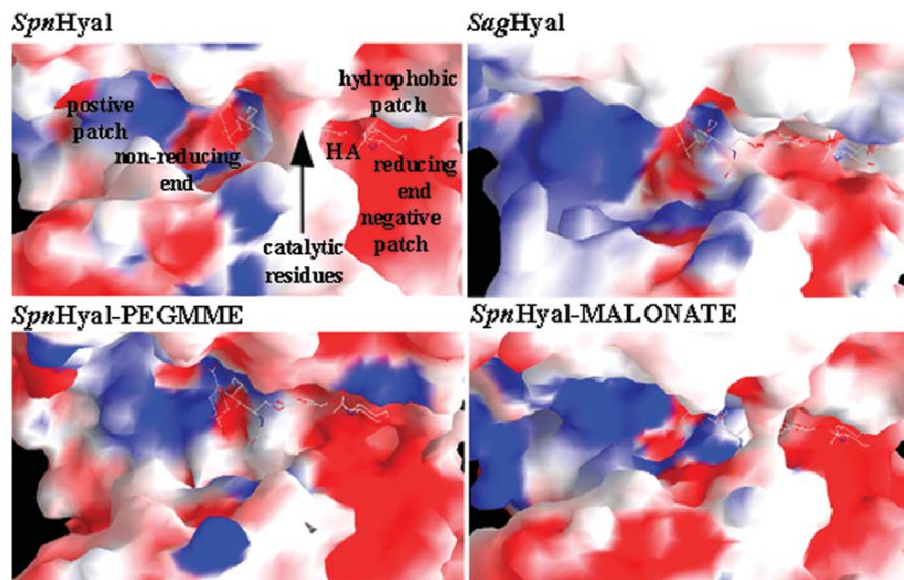
Figure 4.2: Differences in the extent of the opening/closing of the catalytic cleft for different crystal structures discussed in the chapter. Side-by-side view of the distribution of the potential inside the catalytic cleft for representative structures. The location of catalytic, hydrophobic, negative, and positive patches as well as the reducing/non-reducing ends of the HA substrate are labeled. The electrostatic potential magnitude is held constant for all enzyme panels, and is colour-coded as follows: blue: positive; red: negative.

loops from the β-domain, relative to the α-domain, that are involved in the formation of the side of the cleft resulting in opening/closing of the access/entrance to the cleft (Fig. 2.5 right). For the sake of clarity we maintain the same numbering of motions throughout the present work.

Here, we present multi-nanosecond molecular dynamics simulations of *Spn*Hyal-*holo* and *Spn*Hyal-*apo* in explicit solvent. These represent a significant advance on the earlier computational analyses made with the computationally efficient but quantitatively less rigorous CONCOORD program.[22, 23, 223] Our simulations confirm the results from the previous flexibility analysis and reveal the three types of domain motion found before as the principal modes of collective fluctuation as derived from a principal component analysis (PCA). In both molecular dynamics (MD) simulations, *Spn*Hyal-*apo* and *Spn*Hyal-*holo*, the opening/closing mode of the cleft (ii) is the largest amplitude fluctuation. As can be seen in Fig. 4.3, the extent of the opening found in the simulations is in line with the x-ray structures. The most open structures are even more widely opened in the simulation of *Spn*Hyal than the *Sag*Hyal and *Spn*Hyal-PEGMME crystal structures. The opening motion is found to be reversible in the simulations, with the most closed conformations demonstrating the same degree of closure as the x-ray structures of the closed enzyme, *Spn*Hyal. Along this opening mode, the *Sag*Hyal structure, and oth-
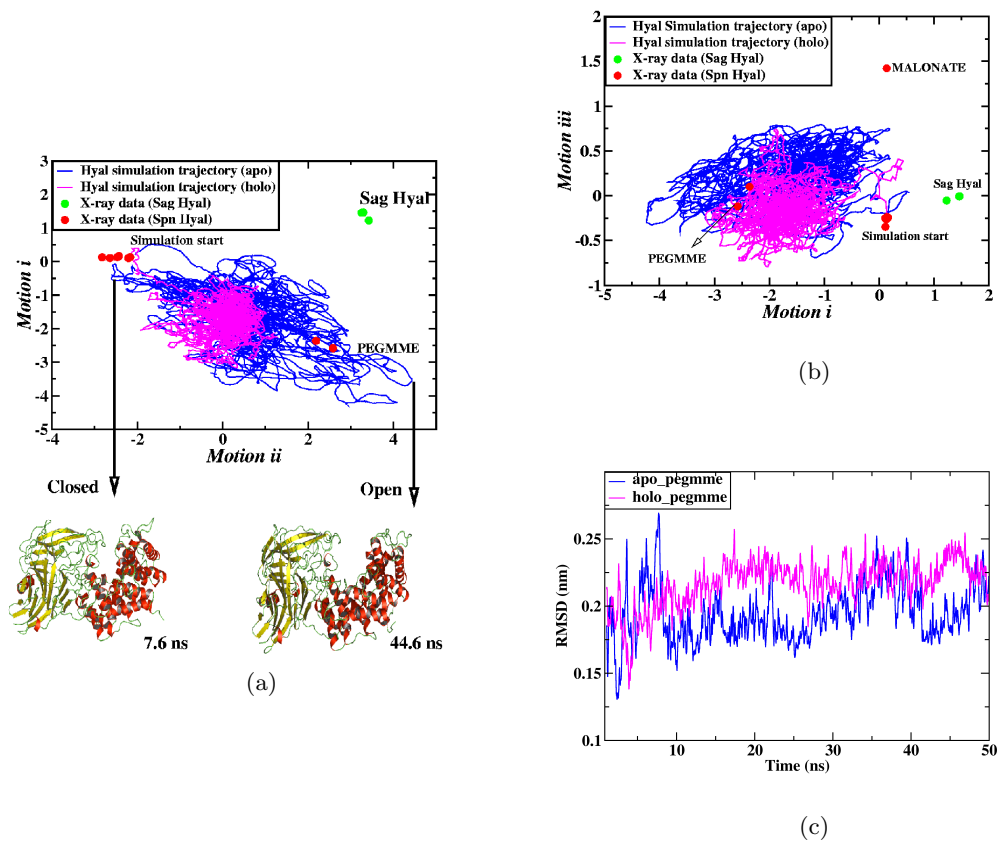
Figure 4.3: MD trajectories (blue and magenta traces) together with x-ray structures of Hyal (circles), projected onto the two principal modes of structural divergence among all bacterial Hyal structures. (a) The graph demonstrates the opening/closing motion of the cleft (motion (ii)) revealed by the principal mode of structural divergence (x-axis). The 50 ns trajectory of the *Spn*Hyal-*apo* simulation (blue) samples this motion in close agreement with the structural data available, whereas the hexasaccharide-bound *Spn*Hyal-*holo* remains in a more closed configuration (magenta). Also along the twisting mode (i), the simulations sample a similar range as the x-ray structures, leaving only the *S. agalactiae* Hyal (Sag.Hyal) x-ray conformations unsampled. The arrows point to two structure snapshots from the simulations, depicting the most open (right) and closed (left) sampled configurations from the simulation. (b) Similar projection of MD trajectories reveals the twisting motion (motion (i); x-axis) and the opening/closing motion of the entrance of the cleft (motion (iii); y-axis). The *Spn*Hyal-*apo* and *Spn*Hyal-*holo* simulations (blue and magenta traces, respectively) sample this motion, again in close agreement with the structural data. (c) RMSD for the two MD trajectories with respect to the PEGMME crystal structure. The blue trace shows the trajectory of the *Spn*Hyal-*apo* simulation and magenta trace shows the trajectory of the *Spn*Hyal-*holo* simulation. RMSD for both the simulations decreases to as low as 1.2 Å showing that the opening mode is significantly sampled.

ers from the same species (green circles in Fig. 4.3), exhibit a slightly higher degree of opening than the *Spn*Hyal-PEGMME structure.

An RMSD (root mean square deviation) analysis was carried out to quantify the structural differences. The RMSD between the crystallographic starting structure of the simulations (PDB code: 1LOH (Table 4.1), the "closed" conformation) and the *Spn*Hyal-PEGMME structure (the "open" conformation) is 2.2 Å. During the simulation of both *Spn*Hyal-*apo* and *Spn*Hyal-*holo* the RMSD, compared to the PEGMME structure, decreases to as little as 1.2 Å (see Fig. 4.3 (c)). This confirms that in both simulations the opening mode is significantly sampled. Interestingly, the MD simulation including a hyaluronan hexasaccharide (*Spn*Hyal-*holo*) does not sample the more open configurations to the same extent as the simulation of *Spn*Hyal-*apo* within the same time span of 50 ns (Fig. 4.3(a)), suggesting that the substrate locks the protein in a more closed conformation. The maximal amplitude of the opening of the cleft of *Spn*Hyal-*apo* is observed to be up to 12.8 Å (measured as the difference between the $C^\alpha - C^\alpha$ separations of Asn231 (in the loop in the $\alpha-$domain) from Gly769 (in the loop in the $\beta-$domain, which contributes to the catalytic cleft) in the most open and the least open trajectory snapshots). In the simulation of *Spn*Hyal-*holo* the extent of the cleft-opening, measured in the same way, is observed only to be up to 6.5 Å (see Fig. 4.3(a) and Table 4.3). This is further quantified by the mean projection along this mode for the *Spn*Hyal-*apo* simulation to 0.64 ± 1.16 (standard deviation) while for the *Spn*Hyal-*holo* simulation it is 0.01 ± 0.51. For comparison, the mean Asn231-Gly769 distances of the *Spn*Hyal-*holo*, *Spn*Hyal-PEGMME, and *Sag*Hyal structures are 14.0 Å, 22.1 Å, and 21.6 Å, respectively.

The second-largest amplitude domain motion observed in the simulations corresponds to the twisting mode (i) from the original flexibility analysis. This mode results in a reciprocal motion of the two domains with respect to the cleft, and might therefore enable an effective translocation of the substrate along the cleft. This presumption is supported by the extent of the twisting motion in the simulations. In the simulation of *Spn*Hyal-*apo* the α-domain is observed to be shifted with respect to the β-domain due to this twisting mode by up to about 11 Å. This extent was measured as the separation between the C^α atoms of Asp340 of the superimposed most twisted and least twisted snapshots from the trajectory. Strikingly, this is also approximately the length of a disaccharide unit of the hyaluronan substrate. Also along this mode, the simulated motions agree closely with the crystallographic structural differences.

Along this twisting mode (i), the *Spn*Hyal-PEGMME structure represents a structural difference to *Spn*Hyal in the opposite direction as the *S. agalactiae* x-ray structures, *Sag*Hyal. However, the MD simulations span the complete range sampled by the x-ray structures, suggesting that this mode also reflects an intrinsic mode of flexibility of the protein. Our results also indicate that, although virtually the whole range of both the opening/closing (ii) and the twisting mode (i) is accessible to *Spn* Hyal, these two modes appear to be coupled during the simulation, leaving the *Sag* Hyal x-ray structures (green circles in Fig. 4.3) unsampled during the simulations. This means the enzyme cannot attain high values for *motion i* and *motion ii* at the same time (see Fig. 4.3). Thus, although the *Spn* Hyal and the *Sag* Hyal x-ray structures are both reflecting the inherent

Eigenvector/motion description	*Spn*Hyal-*apo* (Å)	*Spn*Hyal-*holo* (Å)
Eigenvector (i) (opening and closing of the cleft)	12.8	6.5
Eigenvector (ii) (twisting of the α-domain)	10.9	5.9
Eigenvector (iii) (opening and closing of the access to the cleft)	9.1	5.5
See the text for the details of the measurements.		

Table 4.3: The extent of the dynamic domain motions in *apo* and *holo*-Hyal simulations. For the first eigenvector the distance between the $C^\alpha - C^\alpha$ separations of Asn231 and Gly769 in the most open and least open trajectory snapshots gives the extent of the opening/closing motion of the cleft. The twisting extent (amplitude of the second eigenvector) is measured by the separation between the C^α atoms of Asp340 for the superimposed most twisted and the least twisted snapshots from the trajectory. For the third eigenvector the distance between the $C^\alpha - C^\alpha$ separations of Asn341 and Asn580 was measured to give the extent of the opening/closing of the access to the cleft at the reducing-end.

flexibility of this family of enzymes, part of the structural difference presumably results from sequence differences.

Mode (iii), corresponding to a bending of the two domains that influences the width of the cleft entrance/exit, is the third-largest amplitude mode resulting from the MD simulations. Along this mode, the *Spn*Hyal structure samples an intermediate position and the *Spn*Hyal-PEGMME structure and *Sag*Hyal structures represent a conformation with the entrance opened and the exit closed. The MD simulations demonstrate that this motion is sampled to the same extent in the opposite direction, thereby enabling opening the exit and closing the entrance to the cleft. The amplitude of this motion was measured as the difference between the $C^\alpha - C^\alpha$ separations of Asn341 and Asn580 (since these residues are from the $\alpha-$ and $\beta-$ domain respectively and are near the exit side of the catalytic cleft) in the trajectory snapshots with most open and least open cleft entrances in the same way as for mode (ii). This distance was found to be maximally 9.08 Å in the simulation of *Spn*Hyal-*apo*, whereas for *Spn*Hyal-*holo* the amplitude of this cleft-access mode was found to be maximally 5.47 Å. Thus, also this motion is significantly more explored in the *Spn*Hyal-*apo* structure.

4.5 Discussion

Two novel structures of *Spn*Hyal have been presented. They were obtained using crystallisation conditions different from those hitherto employed. It could be argued that they do not necessarily reflect physiologically relevant conformations due to the

use of PEGMME and MALONATE. However, several factors lead us to believe in the validity and relevance of the structures. Although precipitant can certainly influence the crystallised conformation, it is unlikely that it could "force" the molecule into an energetically unfavourable conformation. In fact, conformational diversity in structures of the same protein has often been correlated with functionally important motions. Thus, the conformations captured in the *Spn*Hyal-PEGMME and *Spn*Hyal-MALONATE structures can reasonably be supposed to be accessible to the enzyme and hence to be of potential importance to its *in vivo* function. In this regard, it should be noted that the pH values at which the structures crystallised, pH 6.5 and 6.6 for the *Spn*Hyal-PEGMME and *Spn*Hyal-MALONATE structures, respectively, are neither far from the pH optimum of the enzyme of ∼ pH 6.0–6.5, nor very different from pH 6.0 at which the enzyme has previously been crystallised. The pneumococcal Hyal is a very resilient enzyme and remains active under a variety of conditions including moderate concentrations of ammonium sulfate (original crystallisation medium for this enzyme), or sodium malonate salts, or PEG, including PEGMME.[219] As expected, however, at high concentrations of these crystallising molecules the enzyme activity is compromised. Regardless, Hyal molecules capable of full activity[22,23] are contained within the crystalline lattice or once such crystals are re-dissolved under standard assay conditions.[213,214]

Importantly, commonly molecular dynamics simulations can reproduce modes of structural variability observed in sets of crystal structures and, indeed, we show that to be the case for Hyal. Finally, the most radically different conformation of *Spn*Hyal described here, *Spn*Hyal-PEGMME, bears a clear resemblance to a previous structure of *Sag*Hyal making less likely the idea that it is an artifact of no physiological relevance.

With the reliability of the new structures established, their implications for our understanding of Hyal flexibility and function have been discussed. Based on MD simulations studies, a cleft opening/closing motion (ii) has been predicted to be essential for catalysis. The new *Spn*Hyal-PEGMME structure now provides the first experimental evidence that directly points to the ability of Hyal to achieve this open conformation predicted earlier. Such an open conformation was previously observed for *Sag*Hyal but the different species origin of that protein left a doubt as to the universality of the flexibility leading to the conformation. The new structure reveals that this cleft-widening motion is an intrinsic mechanistic property of the enzyme, independent of the protein source.

The structural differences between all Hyals may be correlated to their modes of action, as revealed by kinetic analysis with a variety of substrates.[215] Biochemical data demonstrate that Hyal cleavage of HA, its primary substrate, proceeds via an initial endolytic cleavage until the average HA chain decreases to approximately 300 kDa. At this stage another mode of action based on processive, exolytic HA degradation of one disaccharide at a time initiates. When HA chains reach ∼ 100 disaccharides (of about 300 kDa) in length the rapid exolytic and processive activity producing a solely unsaturated disaccharide unit as the end-product of degradation are the primary mode of action.[22,210] Such behaviour is directly related to HA aggregation. Large HA polymers strongly aggregate whereas aggregation decreases with diminishing size of HA. Below 40 kDa (100 disaccharides)[254,255] HA is thought not to be aggregated, resulting

in processive action of Hyal. Bacterial Hyals degrade chondroitin substrates by a purely endolytic process.

With the benefit of the new crystal structures presented here, and the consistent multinanosecond MD simulations reported, it is tempting to speculate on a putative underlying molecular mechanism of processive degradation of substrate(s) by bacterial Hyals. The process putatively consists of the following steps. Once the access to the cleft is open (motion (iii)) and the cleft is widely open (motion (ii)) the substrate is free to bind in the cleft. Upon binding, the entrance closes (motion (iii)) and the cleft width diminishes while the catalytic residues move into their functional positions (motion (ii)). Once the catalytic residues perform catalysis, the residues responsible for product release move in to repel the product (motion (ii)). For further processive exolytic degradation of one substrate disaccharide at a time, the entrance-gate remains closed (motion (iii)) preventing the substrate from escaping the cleft and the cleft widens again (motion (ii)). The substrate-bound enzyme then undergoes the twisting motion (i), by which the substrate is translocated in the direction of its reducing end by $\sim 10\,\text{Å}$, which is also the length of a disaccharide. This would be followed by motion (ii), cleft closure, and catalysis and subsequent cleft opening and product release, and repetition of the cycle until the substrate is fully degraded in the processive manner. In the endolytic degradation mode, both cleaved products would be released from the cleft after the catalysis, a process facilitated by gate opening (motion (iii)) and widening of the cleft (motion (ii)). The type of degradation is substrate specific and depends on the length and aggregation of the substrate.[218] Further experiments complemented by additional simulations (Chapters 5 and 6) are required to study the individual steps in detail to verify the proposed mechanism.

4.6 Conclusions

We have studied the structural basis of the molecular mechanism underlying the cleavage of substrates by bacterial Hyals with molecular dynamics simulations and x-ray crystallography in a collaborative effort. It is not common for enzymes to crystallise under different conditions yielding different structures of the same enzyme. However, in the case of Hyals from two streptococcal species, *Spn* Hyal and *Sag* Hyal, crystals in a number of structural conformations were obtained. These conformations were found to closely agree with the dynamic behaviour of the enzyme as revealed by molecular dynamics simulations and were convincingly associated with the mechanism of the enzyme for processive hyaluronan degradation at the molecular level. Obtaining the new structures allowed the validation of the computational models of enzyme action. The new structures also reveal that the structural differences previously observed for different bacterial Hyals (*Spn* and *Sag*) are not primarily due to sequence differences. Rather, these structures, together with molecular dynamics simulation results, reflect a set of specific domain motions likely to be common to all bacterial Hyals.
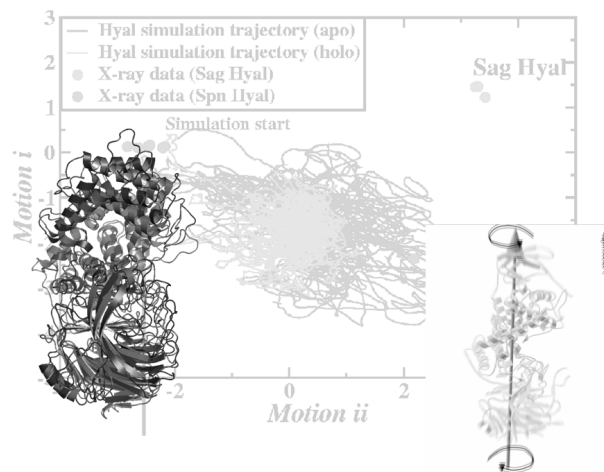
Chapter 5

Flexibility of Hyaluronate Lyase: Part II

Summary

Inspired by the results obtained from the previous chapter, in this chapter we present simulations of *Spn*Hyal together with the HA substrate. Additionally, the two simulations reported in the previous chapter were extended to 100 ns, to see whether there was any change in the domain motions on longer timescales. After a brief introduction (Sec. 5.1) the simulation details are discussed in Sec. 5.2, and in Sec. 5.3 the results emerging from these simulations are discussed. Three loop regions around the cleft have been found to be highly flexible and these may play an important role in the processive mechanism (Sec. 5.3.1). Subsequently, large scale motions of the protein as well as their relation to substrate translocation were investigated. The opening/closing motion of the cleft and the twisting motion of the domains are the largest domain motions in the protein irrespective of the presence or absence of the HA. However, their amplitudes change with the presence of HA and thus corroborating the functional role of these motions in the processivity (Sec. 5.3.2). There was no spontaneous translocation of a complete disaccharide observed. Hence, to investigate the cause for the substrate to remain bound inside the cleft, we first monitored the flexibility of the sugar inside the cleft (Sec. 5.3.3). The relation of the domain motions to the HA-substrate flexibility are discussed in Sec. 5.3.4 and 5.3.5. These observations and the investigations of the interactions of the HA substrate with specific regions and some of the critical residues of the cleft known from the mutation experiments (Sec. 5.3.6 and Sec. 5.3.7) gave some insight into the interplay between the protein domain motions and the processivity of the sugar. Finally, the probable cause of an energy barrier that may be present for the processivity to occur spontaneously is discussed in Sec. 5.3.9. The chapter ends with the collection of all the results in Sec. 5.4.

5.1   Flexibility of *Spn*Hyal

In Chapter 4, it was observed that *Spn*Hyals exhibit flexibility in the form of domain motions. In the case of Hyals from two streptococcal species, *S. pneumoniae* and *S. agalactiae*, crystals in a number of structural conformations are available and these conformations were found to closely agree with the dynamic behaviour of the enzyme as studied by molecular dynamics simulations. Furthermore, this flexibility was experimentally confirmed by two newly resolved structures (PEGMME and MALONATE as discussed in Chapter 4) in the "open" conformation. Obtaining these new structures allowed further validation of the computational models of enzyme action, which we now extend for the further studies.

In the previous Chapter, the flexibility of *Spn*Hyal was hypothesised to be associated with the mechanism of processive hyaluronan degradation at the molecular level. In the proposed model, the flexibility of Hyal was suggested to facilitate processivity of this enzyme by threading the linear polymeric substrate(s) through the enzymatic/catalytic cleft. Against that background it is now interesting to see whether additional simulations including the HA support that mechanism. In this chapter we study the flexibility of the *Spn*Hyals with HA substrates of different lengths. With long timescale simulations ($\sim 100\,\text{ns}$), we investigate the question in detail how the flexibility of the enzyme and the processive translocation of short length HA-substrates are linked. Basically, two questions are addressed: (1) does the flexibility of the enzyme depend on the presence of HA substrates of different lengths, and (2) do we observe any spontaneous movement of HA-substrates of these lengths?

Related to these questions is the question to determine whether two classes of interactions between enzyme and substrate can be distinguished. It was proposed earlier that processive enzymes provide both a specific, tight binding mode that is necessary to optimally position the substrate for catalysis, and an unspecific, weak binding mode for the sliding phase. The second, weak binding regime can be expected to be characterised by a trade-off between strong enough binding to prevent the substrate from dissociating, and a weak enough binding to enable sliding of the substrate between successive rounds of catalysis. We are interested in knowing whether indeed two distinct binding regimes can be distinguished or not, and if yes to study the nature of both binding modes, both structurally and energetically. Since *Spn*Hyals are processive in nature, they remain attached to the HA between the multiple rounds of the catalysis. Thus they should exert a strong binding force on the HA so that the HA does not leave the cleft in the catalytic phase, while they should at the same time exert weak forces or some repulsive mechanism in the sliding phase so that the HA can slide through the cleft easily. We wish to investigate how the *Spn*Hyal achieves this.

No.	Simulation	protein structure	substrate length	time [ns]	Total System size
1.	*Spn*Hyal-*apo*	1loh	—	100 (5)	77,278 (7)
2.	*Spn*Hyal-*holo*	1loh	hexasaccharide	100 (5)	83,542 (10)
3.	*Spn*Hyal-*holo*2	1loh	hexasaccharide	100 (5)	83,542 (10)
4.	HLT2water1	1loh	tetrasaccharide	100 (5)	89,550 (9)
5.	HLT2water2	1loh	tetrasaccharide	85 (5)	89,167 (9)
6.	HLT2water3	1loh structure back mutated to F408Y	tetrasaccharide	95 (5)	89,551 (9)
7.	HLT2water4	1loh structure back mutated to F408Y	tetrasaccharide	40 (5)	89,460 (9)
8.	HLT3water1	1loh	hexasaccharide	100 (7.5)	100,681 (10)
9.	HLT3water2	1loh	hexasaccharide	100 (7.5)	100,682 (10)
10.	HLT3water3	1loh	hexasaccharide	100 (7.5)	100,486 (10)
11.	HLT4water1	1loh	octasaccharide	75 (10)	162,681 (11)
12.	HLT4water2	1loh	octasaccharide	46 (10)	165,341 (11)
13.	HLT4water3	1loh	octasaccharide	70 (10)	165,341 (11)

Table 5.1: List of simulation systems with the simulation time in ns (equilibration time in brackets) and total number of atoms in the system. The number in the brackets of the last column relates to number of counterions (Na^+) used to neutralise the system charge.

5.2 Simulation details

Several free MD ('real time MD') simulations of ligand-bound and ligand free *Spn*Hyal were carried out with the MD software package GROMACS 3.1.4, GROMACS 3.2.1 and GROMACS 3.3.[81–83] The starting structure for all the simulations was the crystal structure with PDB code: 1LOH.pdb.[22] The two simulations discussed in the previous chapter, namely *Spn*Hyal-*apo* and *Spn*Hyal-*holo*, were extended to 100 ns. Another simulation of *Spn*Hyal-*holo* was also carried out for 100 ns but with different initial conditions of coordinates and velocities (coordinates of the previous *Spn*Hyal-*holo* simulation at 10 ns as a starting structure). In the new simulations the HA-substrate was modified such that it resembles the substrate just after the catalytic action has been performed (completion of step 4 in PAD mechanism discussed in Sec. 2.3.2). That way the HA-substrate in the starting structure is now ready for the next catalytic round *i.e.* for the sliding through the cleft, according to the proposed mechanism of processivity.

Simulations including tetra-, hexa- and octasaccharide were carried out. For the simulations with tetrasaccharide-HA as the substrate, the first part of the hexasaccharide (HA1 from Fig. 2.4) was manually removed and the remaining tetrasaccharide part was treated as substrate. Two types of simulations were performed — one with the crystal structure 1loh and the other with wild-type protein. The wild-type protein was back-

mutated to F408Y, since in 1LOH.pdb the catalytic residue Tyr408 was mutated. Since this mutation concerns solely a catalytic residue it presumably does not affect the processivity mechanism (the sliding phase of the HA-substrate). For the simulations with hexasaccharide-HA as the substrate, a disaccharide was modelled at the non-reducing end of the tetrasaccharide part obtained as described above. This resembles the 1loh structure but shifted by one disaccharide unit towards the non-reducing end. The subsequent simulations were performed to see whether in 'free' simulations the HA-substrate would slide through the cleft to yield a 1loh-like structure. Similarly, for two of the simulations (no. 11 and 12) with octasaccharide as substrate the initial structures were also modeled such that octasaccharide is at the start of the next round of catalytic cycle or sliding phase. For simulation no. 13, with the octasaccharide as substrate, the HA was placed like in the 1loh-like structure. This means that the end-product side is occupied and the substrate is now ready for the catalytic action. The details of each simulation system are given in Table 5.1. Each series of simulations identical in structure and number of atoms differs in the initial conditions of coordinates and velocities (in a similar way as *holo* and *holo*2).

The simulation systems were set up with simulation details as discussed in Chapter 4. The equilibration time for simulations was now increased up to 10 ns for the modeled systems to accommodate the changes in the substrate structure. The total simulation time was more than $1\,\mu$s. As before, the coordinates at every picosecond were recorded for analysis.

5.3 Results and discussions

From all the simulations listed in Table 5.1, we first identified the major regions of the protein that exhibit large flexibility during the simulations and thus may play a role in the processivity mechanism (Sec. 5.3.1). Since the main features of the protein flexibility are described as the first three principal modes of motion from the PCA (see Chapter 4), we investigated these modes again for all simulations (Sec. 5.3.2). To investigate the relationship between the sugar motion and the protein flexibility, first the flexibility of the sugar in terms of its root mean square deviation (RMSD) and its motion inside the cleft were monitored (Sec. 5.3.3). To see whether any spontaneous processivity was observed in the free simulations the translocation of the HA-substrate through the cleft was monitored in Sec. 5.3.4. The role of the domain motions of the protein with respect to the motion of the sugar as a measure of the processivity gives an indication of how protein domains affect the motion of the sugar inside the cleft. Finally, specific interactions between the cleft-residues and HA occurring inside the cleft were monitored to investigate in detail the functional role of protein-domain motions in the processivity mechanism. We characterise and rationalise some of the specific known interactions from the protein-sugar complex in Sec. 5.3.6 to Sec. 5.3.8. Based on these interactions, major factors that may be responsible for the HA-substrate to remain attached inside the cleft are validated and proposed.

5.3.1 Equilibration and the flexibility of the *Spn*Hyals

The root mean-square deviation (RMSD) from the crystal structure 1loh (*i.e.* the starting structure of *Spn*Hyal-*holo* simulation) was monitored during the equilibration.
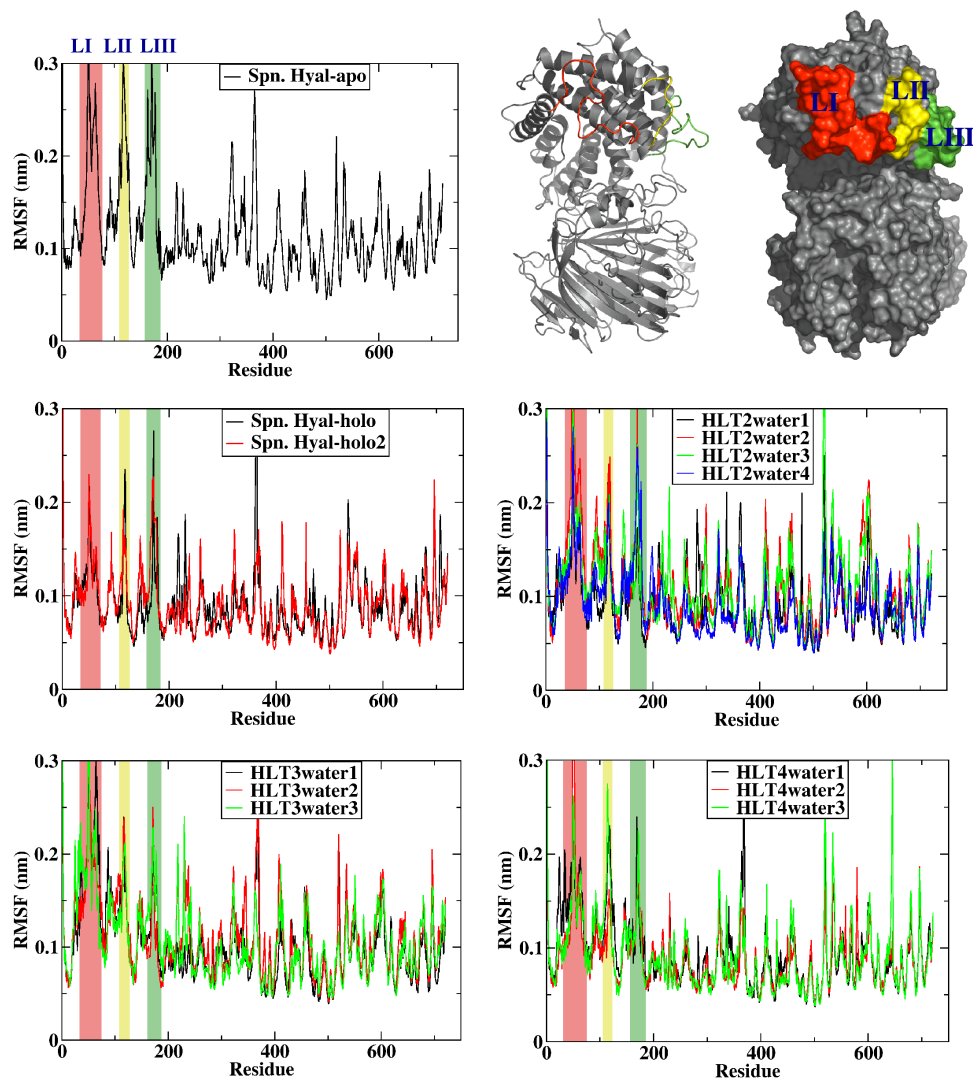
Figure 5.1: RMSF per residue for all the simulations. In the upper right corner the protein is shown in cartoon and in surface representation with the regions of the three loops marked with LI: red, LII: yellow, and LIII: green. The RMSF for these loops are highlighted in the same colours in all the graphs

For all simulations, the RMSD for the protein C^α atoms was found to level off between $1.25\,\text{Å}$ to $2.5\,\text{Å}$ (not shown). The root mean square fluctuations (RMSF) per residue for the *Spn*Hyal-*apo* simulation and all other simulations are shown in Fig. 5.1.

A striking feature is the suppressed RMSF for the three flexible loops indicated as LI, LII, and LIII in the *holo*-simulations as compared to the *apo*-simulation. In the *apo*-simulation the RMSF for these loops is very high, up to 4.0 Å. This fluctuation is lower in all cases in which the sugar is present in the cleft, showing the influence of its presence. On the contrary, the RMSF for the residues 27-43, which form a helix at the entry side of cleft (just preceding the loop LI), increases for most simulations with HA-substrate (simulations 4–13 in Table 5.1) except in *holo*-simulations and simulations with octasaccharide substrate (simulations 2–3 and 12–13 in Table 5.1). In those cases, where the RMSF for this helix is high (sim. no. 4–12), the position of the HA-substrate is such that it is ready for the next round of the catalysis. The RMSF for the residues 338–350, which form a loop area near the entry side, shows a similar behaviour. The loop region from the $\beta-$domain contributing to the catalytic cleft (residues 597–603) also shows increased flexibility in the presence of HA-substrate. The high flexibility for residues 360 to 370 corresponds to the peptide linker between the $\alpha-$ and $\beta-$domain. The high flexibility of the regions around the residues 517 to 524 (in the HLT2water and HLT4water cases) and of residues 634 to 648 (in the HLT4water case) are not of interest here as these regions (loops in the $\beta-$domain) are not near the catalytic cleft. Thus, Overall, protein flexibility seems to be directly affected by the presence of sugar inside the cleft. In the next sections we will study the behaviour of these regions in detail.

5.3.2   PCA and the domain motions in SpnHyals

In the previous chapter, the flexibility of *Spn*Hyal was studied by the analysis of two simulations. Here, we extend this analysis by inspection of additional simulations (see Table 5.1). This is illustrated in the two figures Fig. 5.2, Fig. 5.3 and Table 5.2.

In Fig. 5.2, the 2-D projections for the first three eigenvectors are shown. In each panel A, B, C, and D the graph on the left side shows the projections onto the first two eigenvectors. These graphs demonstrate the opening/closing motion of the cleft on the x-axis and the twisting motion along the y-axis. The green circles in the upper right corner represent the *S. agalactiae* x-ray conformations. All the 13 simulations leave these conformations unsampled, reconfirming the structural divergence among bacterial Hyal structures as was speculated from the earlier flexibility analyses. All simulations in which the HA is in 1loh-like position (both *Spn*Hyal-*holo* simulations (sim. no. 2 and 3, red and green traces from A-left) and the HLT4water3 simulation (sim. no. 13, green trace from D-left)) and one HLT2water simulation (sim. no. 4, black trace in B-left) see Table 5.2) sample the opening/closing motion of the cleft reversibly, but to a lower extent than other simulations some of which sample this motion even up to the newly solved "open" conformations (PEGMME, cyan circles). All these simulations also show the twisting mode significantly sampled. This is also seen from the graphs on the right in each panel where the x-axis gives the extent of the twisting mode.

Figure 5.3 shows the time evolution of the first three eigenvectors for all simulations. For the *Spn*Hyal-*apo* simulation, the first eigenvector, which describes the opening/closing motion of the cleft, is spanned reversibly. The opening at 20 ns repeats again

67
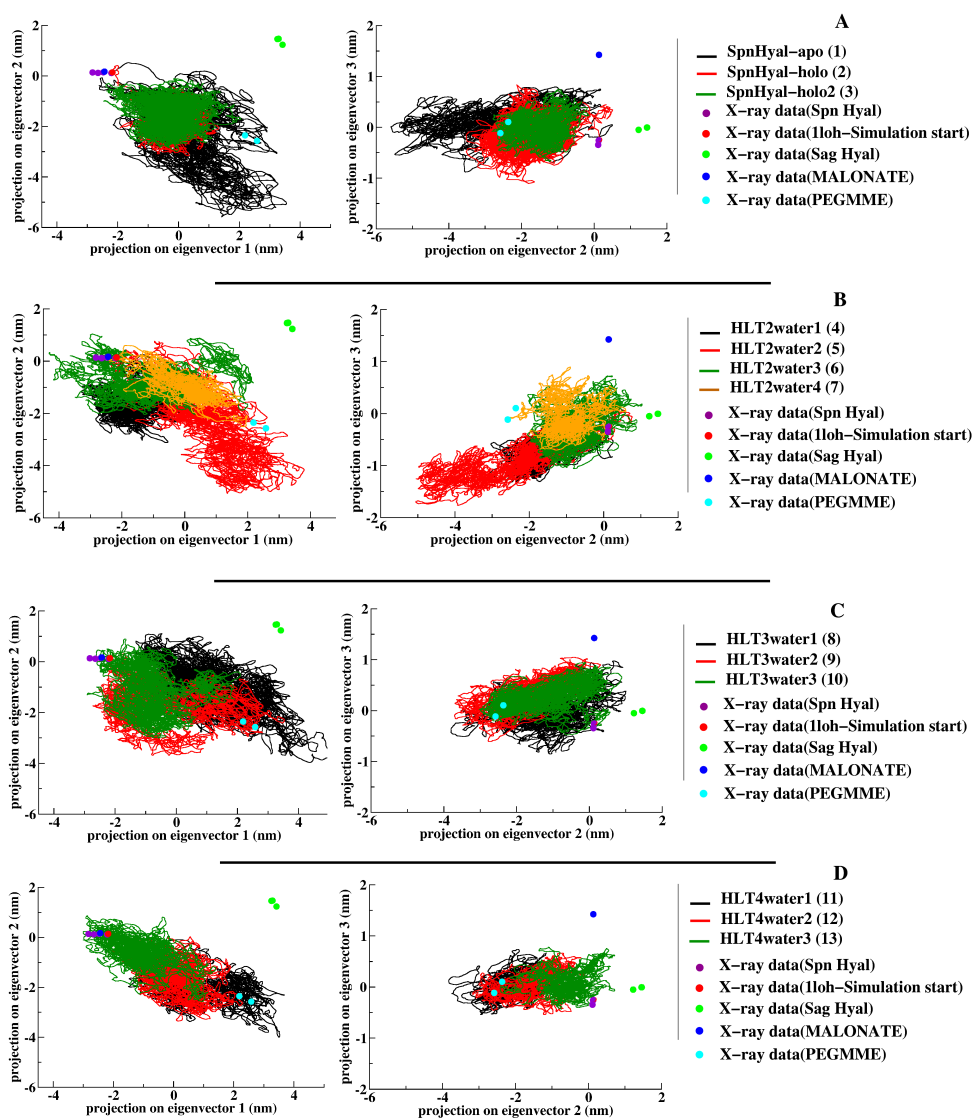
Figure 5.2: 2-D Projections of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. On the left hand side MD trajectories with the available x-ray structures of Hyal (circles) are projected onto the first two principal modes of the PCA. The x-axis gives the extent of the opening/closing motion of the cleft, while the second eigenvector (y-axis) shows the extent of the twisting mode (twisting between α and β domain). On the right hand side panel, the projections on the second and third eigenvector are shown, with the third eigenvector describing the opening/closing motion of the entrance of the cleft (y-axis). (A) Projections of the MD trajectories of the simulations of Hyal-*apo* and two Hyal-*holo* simulations (simulations of 1loh structures). Similar projections of the trajectories of simulations of *Spn*Hyal with the (B) hexasaccharide HA, (C) tetrasaccharide HA, and (D) octasaccharide HA inside the cleft.
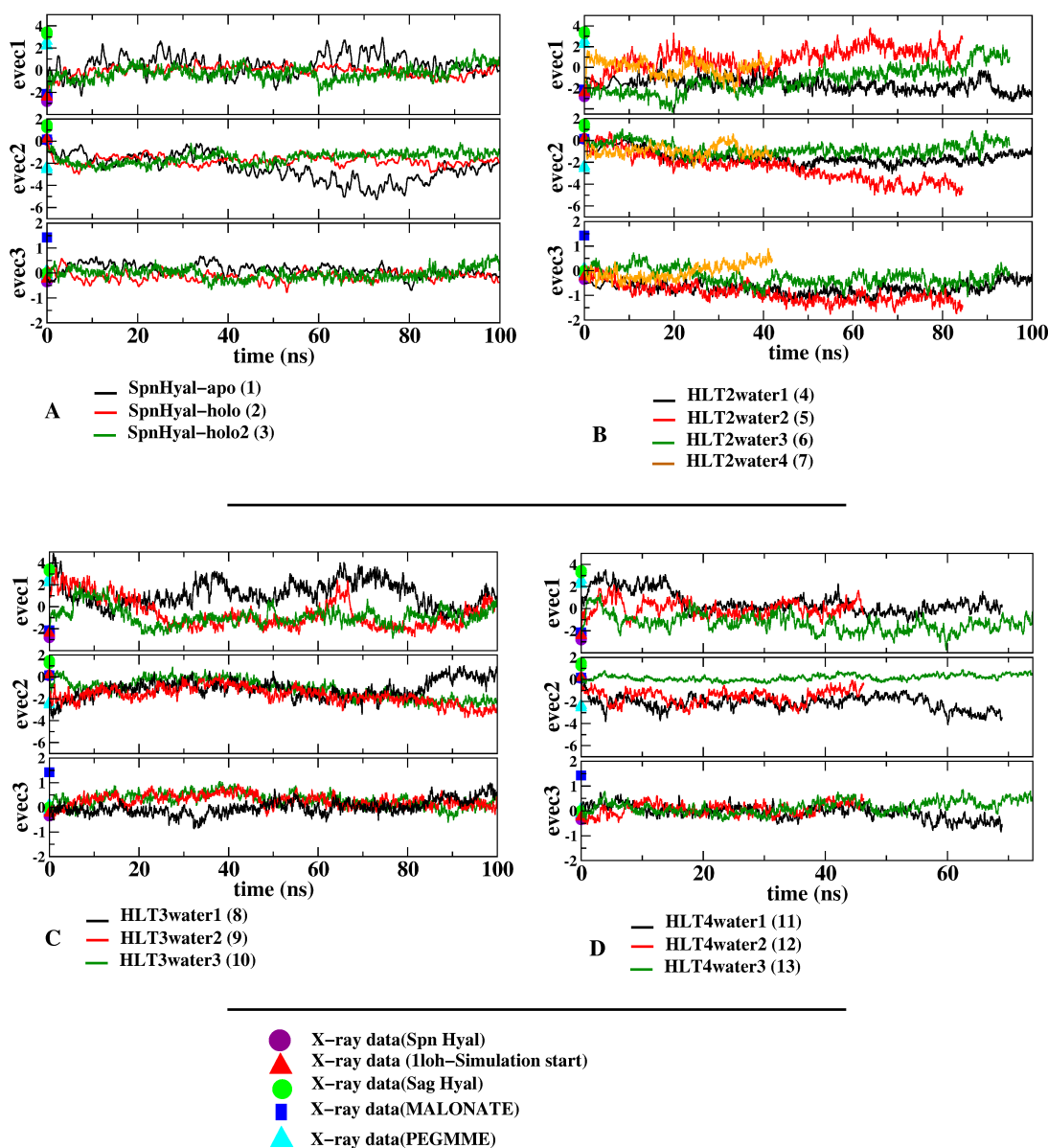
Figure 5.3: Time evolution of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. The available x-ray structures of Hyal are shown as circles and the simulation trajectories are shown as traces. (A) Projections of the MD trajectories of the simulations of Hyal-*apo* and two Hyal-*holo* simulations (1loh structures). Similar projections of the trajectories of simulations of *Spn*Hyal with (B) the hexasaccharide HA, (C) the tetrasaccharide HA, and (D) the octasaccharide HA inside the cleft.

Sim. No.	Eigenvector (i) (opening/closing of the cleft)	Eigenvector (ii) (twisting of the $\alpha-$ domain)	Eigenvector (iii) (opening/closing of the access to the cleft)	Eigenvector (iii) (opening/closing of the exit to the cleft)
1	13.0	11.2	6.3	9.8
2	6.9	6.7	5.9	6.1
3	6.2	4.9	3.9	3.4
4	2.5	4.8	5.7	6.2
5	13.1	9.9	11.4	0.5
6	11.8	4.5	9.5	2.1
7	7.2	2.3	5.1	8.1
8	13.7	4.0	10.0	2.7
9	9.2	3.8	5.8	4.1
10	7.3	5.7	6.2	9.3
11	7.2	4.7	4.4	2.3
12	7.0	7.6	5.2	2.2
13	6.5	5.6	1.6	2.6

Table 5.2: The extent of the dynamic domain motions in all the simulations (same as described in Chapter 4). The order of the simulations is the same as shown in the Table 5.1. For the first eigenvector the distance between the $C^\alpha - C^\alpha$ separations of Asn231 and Gly769 in the most open and least open trajectory snapshots gives the extent of the opening/closing motion of the cleft. The twisting extent (amplitude of the second eigenvector) is measured by the separation between the C^α atoms of Asp340 for the superimposed most twisted and the least twisted snapshots from the trajectory. The last two columns give the extent of the amplitudes for the third eigenvector showing entry-side opening and exit-side opening respectively. For the entry site opening the distance between the $C^\alpha - C^\alpha$ separations of Asp211 and Ser771 is measured while for the exit side opening the distance between the $C^\alpha - C^\alpha$ separations of Asn341 and Asn580 is measured to give the extent of the opening/closing of the access to the cleft.

after about 40–50 ns. This reoccurring opening is also seen in the two of the HLT2water simulations (red and orange traces in B), and one of the HLT3water simulations (black trace in C). Two simulations with hexasaccharide substrate (sim. no. 8 and 9) start from relatively 'open' conformations, but significantly sample the 'closed' conformations and 'open' conformations reversibly as seen from Fig. 5.3 (black and red trace). In HLT4water simulations the protein opens up almost immediately, and then closes for all the three simulations for most of the time. In the simulations with octasaccharide in 1loh-like position, this closing is faster than in the other simulations and to a larger extent. Relating the twisting mode to the cleft opening/closing an anti-correlation can be seen for at least six simulations (*apo* simulation (1), HLT3water (8,9,10) simulations, and two of the HLT2water simulations (5,7) — red and orange traces). The twisting (second eigenvector) is seen to be at the maxima whenever the protein is in the "open conformation" (first eigenvector).

In general, the most open structures from the simulations are even more widely

open than the *Sag*Hyal and the newly solved *Spn*Hyal-PEGMME crystal structures for as many as 6 simulations (sim. no. 1, 5, 6, 8, 9, 11). The extent of the principal modes during the simulations was quantified in a similar manner as in the previous chapter and is summarised in Table 5.2. The maximal amplitude of the opening of the cleft of *Spn*Hyal-*apo* was observed to be up to 13.0 Å (measured as the difference between the $C^\alpha - C^\alpha$ separations of Asn231 and Gly769 in the most open and least open trajectory snapshots). For comparison the distances between these residues in a 'closed' conformation of 1loh and in 'open' conformations of PEGMME (2 structures) were observed to be 14.8 Å, 22.1 Å, and 21.6 Å respectively, showing the maximum amplitude of opening of the cleft in the experimental structures to be 7.3 Å.

Interestingly, both simulations of the *Spn*Hyal-*holo* system (sim. no. 2 and 3) failed to sample the more open configurations to the same extent as the other simulations within the time span of 100 ns, suggesting that the substrate locks the protein in a more closed conformation. Here, the extent of the cleft-opening was observed to be only up to 6.9 Å, and 6.2 Å respectively (see Table 5.2). This observation was further supported by another simulation having 1loh-like structure as a starting point (HLT4water3 simulation (sim. no. 13)), in which the opening was found to be maximally up to only 6.5 Å. All other simulations (except sim. no. 4, black trace, Table 5.2) show varying but generally larger extent along this mode. These results suggest that the decreased flexibility of the protein in the presence of the substrate is due to the HA in the position ready to be cleaved. Thus the *Spn*Hyal-*holo* simulation presented in the previous chapter did not get 'locked-up' in a closed configuration by chance.

The second eigenvector represents a twisting mode of the two domains with respect to the cleft. This mode has therefore been speculated to play an important role in the effective translocation of the substrate along the cleft. The simulated motions along this mode also agree closely with the crystallographic structural differences. In the simulation of *Spn*Hyal-*apo* the α-domain is observed to be shifted by up to about 11 Å with respect to the β-domain due to this twisting mode. This extent was measured as the separation between the C^α atoms of Asp340 of the superimposed most twisted and least twisted snapshots from the trajectory. Strikingly, this is also approximately the length of a disaccharide unit of the hyaluronan substrate. The amplitude of the second eigenvector is seen at its maximum extent in the time span of 60 to 80 ns of the *apo*-simulation (Fig. 5.3 A, black trace).

Only one simulation from the HLT2water system (sim. no. 5) shows the twisting of the $\alpha-$domain up to the same extent (compare red trace in Fig. 5.2 B and Fig. 5.3 B). The simulations for both *holo* proteins do not show as much twisting motion as the *apo*-simulation does. The observation that in all the other 11 simulations the extent of this twisting motion is smaller and that in none of these simulations the substrate is translocated through one complete disaccharide unit supports the role of this twisting mode in the processivity mechanism. For comparison the twisting between PEGMME structures (cyan circles in Fig. 5.2) and 1loh structure was found to be 4.1 Å and 4.0 Å and the twisting between the MALONATE structure and 1loh was found to be 0.4 Å.

The third-largest eigenvector is the third-largest amplitude mode resulting from the MD simulations and corresponds to a bending of the two domains that influences the width of the cleft entrance/exit. Along this mode, the *Spn*Hyal-PEGMME structure and *Sag*Hyal structures represent a conformation with the entrance opened and the exit closed. The MD simulations demonstrate that this motion is sampled to the same extent on both sides of the *Spn*Hyal-PEGMME structure and *Sag*Hyal structures, thereby enabling opening the exit and closing the entrance to the cleft as well as opening of the cleft exit. The amplitude of this motion was measured in both directions. The opening of the cleft entrance was measured as the difference between the $C^{\alpha} - C^{\alpha}$ separations of Asp211 and Ser771 in the trajectory snapshots with most open and least open cleft entrances in the same way as for the first eigen-mode whereas the opening of the cleft exit was measured as the difference between the $C^{\alpha} - C^{\alpha}$ separations of Asn341 and Asn580 for the same snapshots. The opening of the cleft-entrance was found to be at maximum when the complete length of the HA-substrate was inside the cleft (*i.e.* for tetrasaccharide and hexasaccharide). Since the length of the cleft is approximately long enough to fit only 3 disaccharide units of HA (a hexasaccharide substrate) the last disaccharide unit of HA in HLT4water systems lies outside the cleft. In those cases (sim. no. 11 & 12) the opening of the cleft access was found to be minimum for the systems in which HA is ready for the sliding. For the simulations in which the HA position is 1loh-like (sim. no. 2, 3 and 13) this opening was found to be even lower. These observations may elucidate the binding mechanism of *Spn*Hyal to HA during the catalysis process. The opening/closing of the cleft-exit was not found to be correlated to the closing/opening of the entrance site. The cleft-exit was found to be maximally opened (9.8 Å) in the simulation of *Spn*Hyal-*apo*, whereas for the rest of the simulations it was seen to vary from 0.5 Å (sim. no. 5) to 9.3 Å (sim. no. 10). Thus, this motion is also significantly more explored in the *Spn*Hyal-*apo* simulation compared to the simulations of hyals with HA substrate. Interestingly for the last simulation (sim. no. 13) where the octasaccharide substrate is in 1loh-like position, the opening of the cleft access is minimal. Comparing (a) the opening of the cleft in case of 2 *holo* simulations and 3 simulations where hexasaccharide is ready for the sliding, and (b) the opening of the cleft in case of simulation of octasaccharide in 1loh-like position (sim. no. 13) and other two simulations with octasaccharide substrate ready for the sliding phase, it is seen that when the substrate is ready for the sliding phase, the cleft access opens. Once the substrate is in position for the catalytic action, the cleft access closes. This indicates a putative functional role of the third eigenvector as helping in binding the substrate after the sliding action is over.

Taken together, the simulations presented in this chapter thus confirm the domain motions observed in the previous chapter. The fact that such a flexibility of the enzyme is seen in all these different configurations strongly indicates that the protein domain motion plays an important role in the processivity mechanism. Additionally, there appears to be a strong correlation between the protein domain motions and the presence of a sugar substrate/product (*holo* configuration). This correlation will be examined in more detail in the next sections.

5.3.3 Flexibility of the HA inside the cleft

As a next step towards understanding the processivity mechanism, we first zoom in the picture to analyse the interactions of sugar with some specific regions inside the cleft. First, to see whether any spontaneous processivity was observed in our free simulations, and also to obtain a measure of sugar flexibility inside the cleft, the positions of the sugar were monitored with respect to the cleft. The resulting distances are shown in Fig. 5.4 for the four cases as discussed. On the x-axis the distance between the center of mass (COM) of the catalytic cleft of the enzyme and the COM of the sugar is plotted whereas on the y-axis the distance between the COM of the next disaccharide to be cleaved and the COM of the catalytic residues in the cleft is measured. For one complete cycle of the processivity this distance should decrease by about 1 nm, which is the length of the HA-disaccharide unit. The distances on the x-axis are scaled to zero such that the starting distance will be at zero for all cases. The movement on the x-axis can be correlated to the overall adjustment or positioning of the sugar inside the cleft *i.e.* the flexibility of the sugar inside the cleft. In all cases the sugar did not show complete disaccharide unit movement along the y-axis towards zero, indicating that processivity was not observed for these cases in at least 100 ns timescale. However, all simulations show the flexibility of sugar in terms of movement along the x-axis to varied extent. In fact in sim. no. 8 and sim. no. 6 the sugar is very flexible and even though the sugar does not leave the protein completely, it shows movement in out-of-plane direction. In sim. no. 8 this can be attributed to the initial condition of the cleft which is relatively opened and for sim. no. 6 it can be attributed to the short substrate, which can move inside the cleft freely. For the simulations of *Spn*Hyal with the tetrasaccharide substrate inside the cleft, the sugar remains stable (Fig. 5.4 B, black, red and orange curves). For the longer substrates (hexasaccharide and octasaccharide) the flexibility of the sugar varies and this may be attributed to the adjustment of the sugar for proper positioning as can be seen from the snapshots (inset in Fig. 5.4 C and D). For the simulations 2 and 3 the sugar remains 'locked up' in its initial position showing not much movement inside the cleft (upper panel of the Fig. 5.4 A).

The flexibility of the sugar inside the cleft as seen from Fig. 5.4 was further supported by the RMSD of the sugar inside the cleft. The RMSD of the first four rings of the sugar in each case was computed with respect to that in the crystal structure 1loh. This allowed direct comparison of all cases discussed above. Fig. 5.5 shows these RMSDs together. For the simulations where the sugar is in 1loh-like position (sim. no. 2, 3 and 13, black, gray and brown traces), the RMSD of the sugar remained stable and lower than all other simulations. Interestingly, RMSDs for the sim. no. 9 and 10 in which sugar was in a position ready for the processivity cycle (orange and magenta traces) started initially from high values and then slowly decreased at around 50–60 ns. The RMSDs for all the other simulations remained stable for the duration of the simulations, and did not decrease. The red trace in Fig. 5.5 where the sugar was ready for the processivity or sliding phase, but with the cleft in a slightly 'open' conformation showed the highest RMSD.
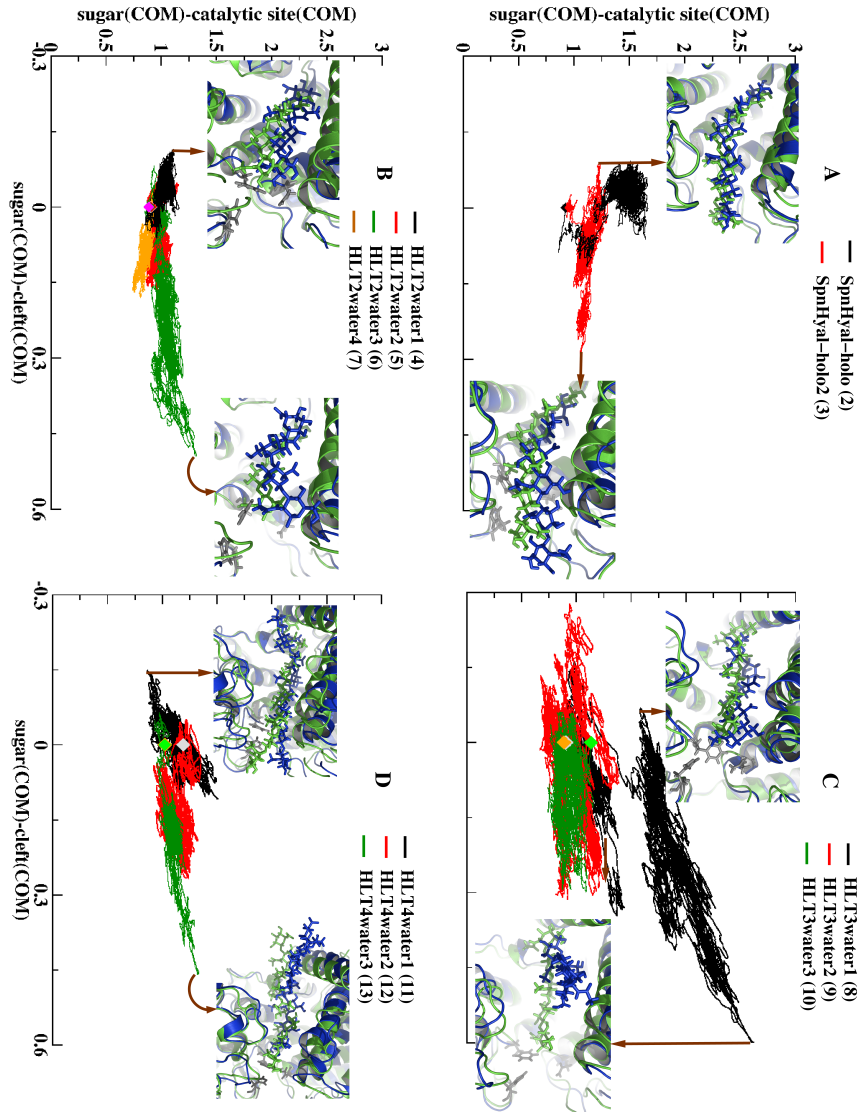
Figure 5.4: Motion of the sugar inside the cleft. On the x-axis the distance between the center of mass of the catalytic cleft of the enzyme and the center of the sugar is plotted. This distance is scaled for each case so that the starting point is zero on x-axis (as shown by the circles). On the y-axis the distance between the center of mass of the next disaccharide to be cleaved and the center of mass of the catalytic residues in the cleft is measured. For one complete cycle of the processivity this distance should therefore decrease by about 1 nm, which is the length of the HA-disaccharide unit. Pictures show the position of the sugar inside the cleft for the particular states shown by the arrows. The green cartoon and stick representation is for the starting configuration while the blue cartoon and stick representation shows the state of the complex at the position indicated by the arrow.
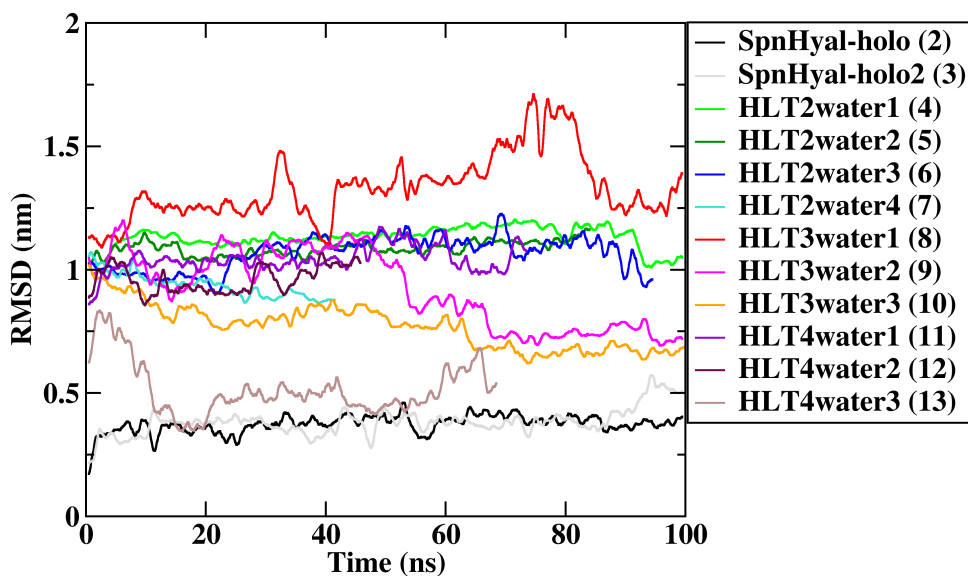
Figure 5.5: Flexibility of the first four rings of HA inside the cleft. RMSDs are calculated with respect to the first four rings of HA in the crystal structure 1loh, which is used as the starting structure for all the simulations in this work. Black, gray and brown traces are the RMSDs for the simulations where the HA is in 1loh-like position (sim. no. 2, 3 and 13). They remain low for all the duration of the simulations. Orange and magenta traces start from the high value where the sugar is ready for the processivity cycle (sim. no. 9 and 10) and slowly decrease to lower values consistently. For all the other simulations, the RMSDs remain stable at high values (except the red trace, sim. no. 8 where the cleft of the protein is relatively open at the beginning than in other simulations)

5.3.4   Processive motion of the sugar

The flexibility of the sugar inside the cleft is shown in Fig. 5.4. Interestingly, the two simulations with hexasaccharide substrates inside the cleft, at a position in which the sugar is ready for the processivity cycle (sim. no. 8–10), showed large motions of the sugar.

To further investigate whether the sugar in each of the above cases has moved towards a reactive (1loh-like) configuration, thus confering the 'real' processive sliding, we projected the simulation trajectory on the first eigenvector obtained from the PCA carried out on a set of two structures: the 1loh crystal structure and the starting structure of HLT3water1 simulation. The first eigenvector of this PCA reflects the effective translocation of the sugar towards a 1loh-like structure. As can be seen from Fig. 5.6 no significant motion of the sugar was observed for the simulations except in the two HLT3water simulations (magenta and orange traces in the uppermost panel A). Both *Spn*Hyal-*holo*
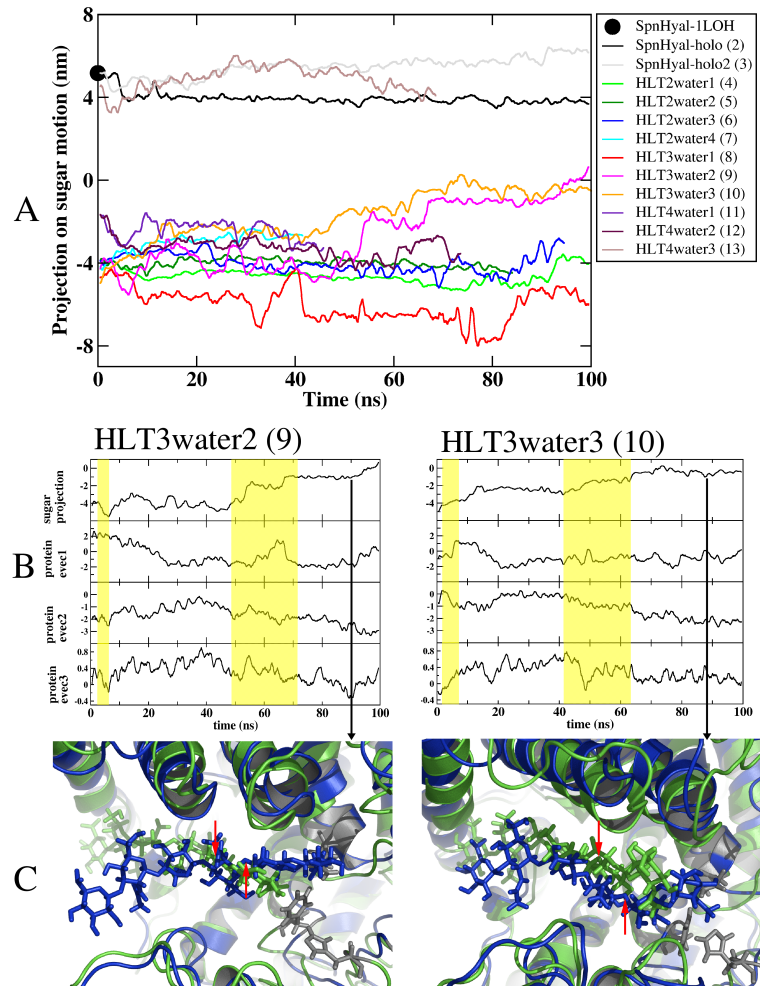
Figure 5.6: Motion of the sugar inside the cleft resulting from the PCA performed on the sugar. Panel A (top) shows the motion of the sugar for all the 12 simulations with HA substrate present inside the cleft. The black circle shows the position of the sugar in the *Spn*Hyal structure. In an idealised processive cycle the simulations ready for the sliding phase (sim. no. 4 to 12) should move the distance of one disaccharide unit and attain the 1loh-like structure (move towards the black circle). Panel B shows a detailed picture of this sugar motion along with three protein domain motions for the simulations 9 and 10, where the sugar shows some processivity in the expected direction. Regions of the sugar motions with respect to the decrease in the second eigenvector value are highlighted in yellow. Panel C shows the overlays of the initial structure of simulations (in green) and the structure at maximum sugar translocation (in blue) for each of the cases. The catalytic position is shown by the residues in grey stick representation. The red arrows show the position of the next glycosidic $\beta - 1, 4$ oxygen linkage that would be cleaved. The arrows pointing down show the initial position while arrows pointing upwards show the final position depicting the movement in the processive direction.

simulations show remarkable stability in the sugar position. The sim. no. 13 where octasaccharide was modelled similar to a 1loh-like position so that it is ready for the catalytic action, also shows stability like both *Spn*Hyal-*holo* simulations. These three simulations indicate that the sugar does not show processivity from the 1loh-like position or the processivity takes place after the catalysis is completed in the step of disaccharide. In all the simulations with tetra- hexa- or octasaccharide substrate except for sim. no. 9 and 10 (magenta and orange traces in the uppermost panel) the sugar remains stable at its original position. Surprisingly in sim. no. 8 (HLT3water1 simulation, red trace in upper panel A), the sugar moves in the non-processive or backward direction. In this simulation, the cleft of the protein was relatively open at the start of the simulations. Interestingly, the two cases of HLT3water (where the hexasaccharide is ready for the sliding phase) spontaneously show almost half way progress towards the 1loh-like position.

Next, we inspect the role of protein domain motions in relation to the processivity for simulations no. 9 and 10. Interestingly, panel B in Fig. 5.6 clearly shows the correlation of the first two eigenvectors emerging from the PCA of the protein domain motions with the translocation of the sugar. The twisting mode or the second eigenvector is apparently connected with the translocation because whenever the projection of this eigenvector decreases ($\alpha-$domain twisting towards exit side of the cleft) the sugar is seen to be moving forward towards the 1loh-like position. The increase in the projection to the first eigenvector or the opening of the cleft seems to be a consequence rather than a cause of sugar translocation. Panel C shows the snapshots of the system at the maximum sugar translocation. In both snapshots, protein and the sugar in initial configuration are shown in green cartoon and stick representation respectively, while the structures at the maximum translocated sugar are depicted in blue cartoon and stick representation. The catalytic residues are shown in gray colour in stick representation. The glycosidic oxygen of the second sugar ring has to come close to these catalytic residues in order to complete one processive cycle. The position of this oxygen is indicated with red arrows in each picture of panel C. From these snapshots it can be seen that the sugar in simulation 9 and 10 has moved only through one sugar ring (*i.e.* about halfway) towards the 1loh-like position. However, Fig. 5.4 also shows that the distance between the next-disaccharide and the catalytic center does not decrease significantly. This shows that this sugar has not properly oriented or positioned yet during the 100 ns of simulation.

5.3.5 Role of protein dynamics in substrate binding

Does the substrate binding affinity, and/or sliding speed depend on protein conformation? In other words is the dynamics of the protein actively involved in the processive mechanism, or is the structural framework offered by the enzyme sufficient for the processive degradation of its substrate?

We now try to correlate the specific types of fluctuations in the protein structure to specific aspects of the processive mechanism in the observations made so far to address above question. It is worthwhile to correlate the highest RMSD in case of sim. no. 8

77

(red trace in Fig. 5.5) and the backward propagation of the sugar along the processivity direction in that case. This movement is seen to be consistent from the beginning of the simulation on (note that the cleft of the protein is open at the start of the simulation). If this simulation is compared with the other two simulations that include a hexasaccharide substrate (sim. no. 9 and 10), and all the other simulations in which no processivity is observed, a striking correlation pattern of the domain motions of the protein with the sugar motion/flexibility emerges. First, this for the first time indicates that either the opening of the cleft or the twisting of the domain motion alone is not sufficient for the complete processivity. Rather sugar translocation appears unfavourable if not assisted by an other domain motion of the protein. In other words this may indicate that in a successful processivity cycle, both motions — twisting between the domains and opening of the cleft — are playing a coupled role. Second, even if the opening of the cleft is seen as a consequence rather than a cause of sugar translocation in sim. no. 9 and 10, it appears a crucial domain motion in the processivity cycle. This seems to be consistent because, as seen from Fig. 5.3 C (black trace), the protein in this case does not show the 'closing' of the cleft to the same extent as in sim. no. 9 and 10 (red and green traces in Fig. 5.3 C), while its twisting motion is similar to those in sim. no. 9 and 10 (at least for the first 80 ns). Hence, if only the twisting motion is important for the processivity, this simulation should also have shown spontaneous processivity in the expected direction. The fact that the protein is relatively open at the start of the simulation thus appears to drastically influence the processivity. Third, there is another piece of evidence to support the mutual interplay of the first two eigen-motions of the protein. Because of the helical nature of HA, it is apparent that the sugar has to reorient itself for the next catalytic process when the glycosidic oxygen is positioned properly relative to the catalytic residues (Fig. 2.4). It is not yet clear whether it can reorient after the sliding phase (slide–and–flip process) or it reorients during sliding (screw–like motion), but it is clear that the opening motion of the cleft is necessary for this reorientation. Since opening of the cleft presumably facilitates the positioning/orientation of the sugar inside the cleft by providing enough space for the sugar, one plausible scenario may be that while the sugar is guided through the cleft with the help of the twisting motion, it orients itself whenever the cleft permits it to do so. During the twisting the cleft tries to remain 'closed' so that sugar does not leave the cleft, but is tightly bound to the cleft. In contrast, when the cleft opens, the sugar tries to reorient itself without sliding. Finally, from the above observations, it may be speculated that this translocation process is happening at much longer timescales (typically more than 100 ns) and poses a main bottleneck of the complete processive cycle as shown in Fig. 2.4.

5.3.6 Interaction of the HA with the positive and aromatic patch of the cleft

We now zoom in to the emerging picture even more to investigate the different energetics and interactions between the sugar and the cleft that may be involved in the processive translocation of the sugar inside the cleft.

For this purpose we investigated particular interactions of the HA inside the cleft for four cases: both *Spn*Hyal-*holo* simulations in which the sugar motion in the processive direction was found to be minimal (Fig. 5.6) and also its flexibility was very low (Fig. 5.5) and the two HLT3water simulations in which we got significant processivity or sliding of the sugar in the expected direction (sim. no. 9 and 10) as observed in the previous section (Fig. 5.6). We chose these four cases for this analysis because the length of the sugar in each case is the same (hexasaccharide) such that these simulations can be directly compared to each other. The simulations differ only in the positioning of the sugar inside the cleft. In the former case (sim. no. 2 and 3) the sugar is ready for catalysis (sliding phase presumably completed) while in the latter cases (sim. no. 9 and 10) the sugar is ready for the processivity or sliding phase. Direct comparison of these simulations may thus give some insight into how the protein-sugar interaction changes at different stages of the complete processive cycle. As discussed in Sec. 2.3.2 the binding cleft of the *Spn*Hyal mainly consists of three regions: (a) a largely positive patch, (b) an aromatic patch near the catalytic residues and (c) a negative patch near the product site. These are shown again in Fig. 5.7. One of the striking features is the charge complementarity of the binding cleft of the enzyme and the HA substrate. The positive patch was proposed earlier to provide an increased affinity for the negatively charged HA substrate. Previously, the aromatic patch (Trp291, Trp292 and Phe343) has been proposed to provide on one hand the basis for a precise positioning of the substrate in the binding cleft,[216] and on the other hand to contribute to the substrate sliding mechanism. We now investigate the roles of these two patches in the processivity for the four chosen simulations.

The short range interactions between the sugar and these regions for the above mentioned four cases are shown in Fig. 5.7. In the topmost panel (A) the initial position of the sugar inside the cleft for each set of simulation is shown. The cleft is shown in surface representation together with the interactive regions of interest (here positive and aromatic patches) are shown. The sugar is shown in space filling representation, and positions of the three disaccharide units D1 (first two rings), D2 and D3 (last four rings) are indicated. In each case D1 would be the next disaccharide unit to be cleaved. In the left picture, the *holo* configuration (sim. no. 2 and 3) shows that sugar is ready for the catalytic phase, *i.e.* the first disaccharide at the reducing end (D1) is ready to be cleaved. The glycosidic oxygen linking D1 and D2 is positioned near catalytic residues (not shown). In the right picture (sim. no. 9 and 10) the sugar is in the sliding/processivity phase and has to move by about one disaccharide unit (~ 1 nm) such that the glycosidic oxygen linking D1 and D2 is positioned at the catalytic residues in the 1loh-like position (or similar to the left picture).

Short range energies of the interaction of the positive patch with the sugar are shown in panel B. This energy is decomposed in two parts – interaction with the first two rings (D1) and interaction with the last four rings (D2 and D3). For *holo* simulations this interaction with the sugar as a whole is strongly attractive (black and red trace for sim. no. 2 and green and blue trace for sim. no. 3) compared to those in HLT3water simulations (black and red trace for sim. no. 9 and green and blue trace for sim. no. 10).
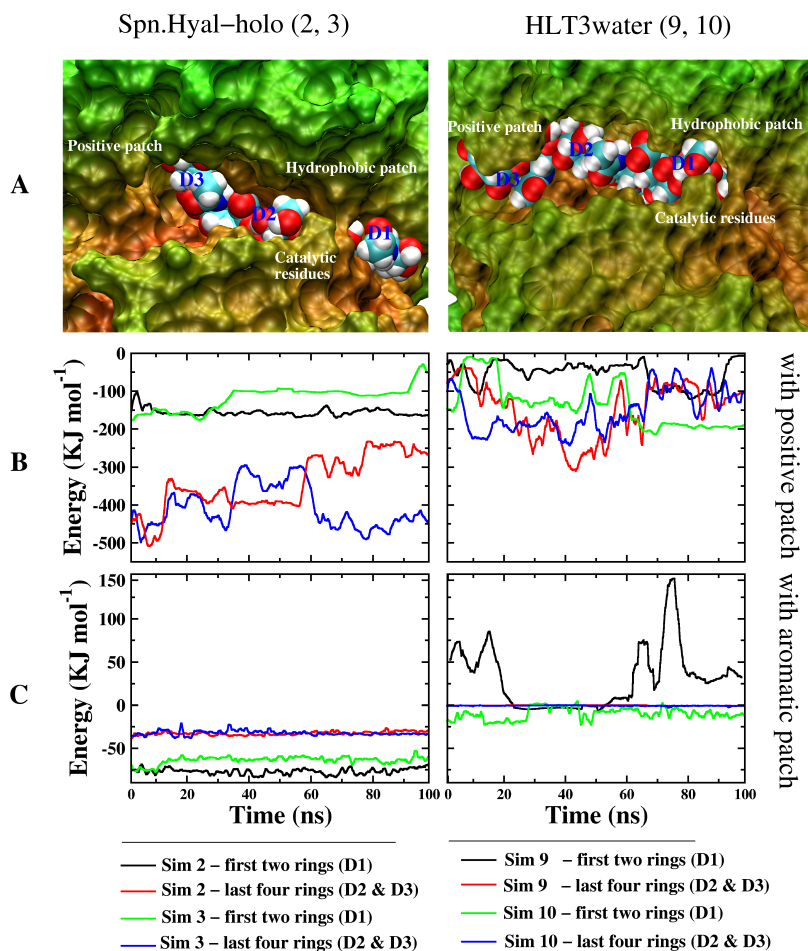
Figure 5.7: A: Surface representation of the interaction regions in the cleft of *Spn*Hyal along with the HA sugar (space filling representation for three disaccharides D1, D2, and D3). A major part of the cleft at the non-reducing end (near D2 and D3) is highly positively charged. The positions of hydrophobic and catalytic residues are also shown. The left picture shows the initial position of the sugar for the *holo* simulations in which the sugar is ready for the catalysis phase whereas the picture on the right shows the sugar ready for the sliding phase (shifted by one disaccharide unit towards the non-reducing end). B: Short range energies of the interactions of the positive patch with first two rings (D1) and the last four rings (D2 and D3) for *holo* (on the left) as well as HLT3water (on the right) simulations. The strong attractive interactions with the substrate (D2 and D3) provide tight binding for the sugar in the catalytic phase (left, red and blue traces) while in the sliding phase, even though the sugar is positioned closer towards the positive patch, a much weaker attraction between cleft and substrate is observed, presumably provided due to the cleft opening. C: Similar interaction energies with the hydrophobic/aromatic patch of the cleft show that for the catalytic phase (left), the non-polar interactions are very stable. The dynamic non-polar interactions in the sliding phase correlate well with the sugar motion and the protein domain motions from Fig. 5.6.

The interaction with D1 is less attractive but stable compared to that with the rest of the substrate (D2 and D3). Similarly, panel C shows the interactions of the hydrophobic/aromatic patch with sugar in two parts (with D1 and with D2 and D3)) for each of the simulations.

In both the *holo* simulations the interactions of D1 with positive patch as well as aromatic patch were observed to be very stable, showing that D1 is 'locked' in the catalytic position. Also, the interaction energies of the last four rings with the aromatic patch for these simulations are very small, mainly due to the larger distance from the aromatic patch. This stability of the energies was also reflected in the mobility of the sugar (See Fig. 5.6) and in the fluctuations of the first 4 rings of the sugar inside the cleft (See Fig. 5.5). This indicates that the fluctuations in the interaction energies with the positive patch in these simulations are largely due to the contribution from the end part of the sugar in these cases (last two rings at the non-reducing end or D3). Except for D3, locking of both the disaccharides around the catalytic patch and aromatic patch can be attributed to 'not opening' of the protein cleft during these simulations. This is further supported by the energy-evolutions in sim. no. 9 and 10. In these simulations, where D1 is not at the catalytic site, the whole sugar is showing dynamic (varying) interactions of the sugar with both positive patch and aromatic patch. As the protein closes, the interaction energy of the last 4 rings with the positive patch becomes more and more attractive and when the cleft is 'closed' (during $\sim 30 - 60$ ns) the interaction with the positive patch is the strongest (most attractive). During this period, the translocation of the sugar inside the cleft is also observed to be minimal (see Fig. 5.6). The interactions with the hydrophobic patch in this period are almost negligible during this period. While in the beginning (first 20 ns) and after about 50 ns the interactions with the aromatic patch were also observed to be stronger, thus supporting the proposed role of the protein dynamics in the processivity mechanism.

Overall, the stronger attractive interactions of both the patches with the sugar in sim. no. 2 and 3 suggest that sugar in 1loh-like position is in favourable position compared to the simulations 9 and 10. In contrast, for the sugar in the processive phase overall weaker attractions (or even repulsive aromatic interactions) indicate an unfavourable position of the sugar inside the cleft. These observations support the hypothesis of strong binding during the catalytic process and weak binding during the processive or substrate-driving phase. However, the fact that the sugar did not show any spontaneous processivity during 100 ns for the simulations 9 and 10 may indicate a large energy barrier that has to be crossed in the substrate-sliding process.

Thus the protein cleft opening/closing motion clearly/strongly alters/affects the attractive interaction of the sugar substrate with the cleft. The disaccharide units around the catalytic site (D1 and D2) are stabilised by both strong attractive interactions with positive patch and by the non-polar interactions with hydrophobic residues. The strong attractive interactions with the substrate (D2 and D3) keep the sugar tightly bound when it is properly oriented for the catalytic phase (left, red and blue traces) while when the sugar is in the processivity/sliding phase the overall attraction for the substrate is lowered even though the sugar is positioned closer towards the positive patch.
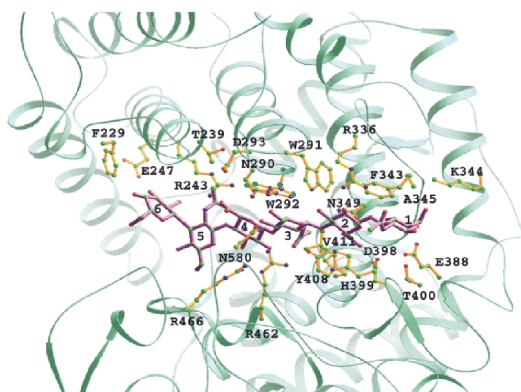
(A) Residues in the spnHL–HA interface

Hexasaccharide sugar rings[a]	Residues in spnHL
Sugar 1	Arg336, Phe343, Lys344, Ala345, Asn349, Glu388, Thr400
Sugar 2	Trp291, Arg336, Phe343, Asn349, Asp398, His399, Tyr408
Sugar 3	Asn290, Trp291, Trp292, Asn349, Tyr408, Val411, Arg462, Asn580
Sugar 4	Arg243, Trp292, Asp293, Arg466
Sugar 5	Thr239, Arg243, Arg466, Asn468, Ser469
Sugar 6	Phe229, Glu247

(B) Selected distances in the spnHL–HA interface

Atoms in HA[b]	Atoms in spnHL	Distance (Å)
C25	His399 NE2	2.46
O22	Arg336 NH1	2.64
O2E	Asn349 OD1	2.06
O2F	Asn349 OD1	2.73
O23	Trp291 NE1	2.38
O24	Tyr408 OH	2.67
O25	Asn349 OD1	3.08
O34	Asn290 ND2	3.16
O36	Asn290 OD1	2.79
O36	Asn580 ND2	3.10
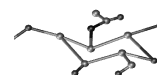
[a]Sugars 1, 3 and 5 are 2-acetamino-2-deoxy-β-D-glucose residues. Sugars 2, 4 and 6 are β-D-glucuronic residues.
[b]The first digit stands for the ring number and the second digit is the conventional atom numbering for sugars.

Figure 5.8: *Spn*Hyal-HA interface. On the left the figure shows the HA-binding site in the *Spn*Hyal structure. The hexasaccharide segment of HA (labeled 1-6) is coloured in purple. The table on the right shows major interface contacts between specific residues from the active site of the enzyme and the atoms of the HA substrate. (Figure and table adopted from S. Li et al.[211]).

Lower energies correlate with the opening of the cleft so that the sugar is relatively weakly bound thus allowing easy translocation but not leaving of the cleft. The interactions with the aromatic patch are also stronger in this period to guide the sugar in the processive direction. These interactions are sometimes even repulsive hinting towards an energy barrier in the translocation phase.

5.3.7 Characterisation of important known residues

In the previous section we observed the role of the positive and aromatic patch in the cleft to provide suitable binding and sliding of the sugar inside the cleft. We now decompose the interactions to the individual residue level to find the important residues that contribute the most to these interactions. To this end we investigated the simulation data for the roles of specific residues that are known from mutant experiments[211] to be important in processivity in one way or another. *Spn*Hyal mutants R243V, N349A, H399A, Y408F (a crystal structure that is also used in all these simulations), N580G, W292A, F343V, W291A/W292A, W291A/W292A/F343V have been previously produced and characterised experimentally. Asn349, His399 and Tyr408 are known to be directly involved in the catalysis since their mutation results in inactivation of the enzyme. Mutants R243V and N580G also show modified activity when compared to the wildtype. A close up of the active center of the *Spn*Hyal along with the hexasaccharide substrate is shown in Fig. 5.8 in which aforementioned residues are shown. Also given are several residues along the cleft in close contact with the hexasaccharide substrate that seem to be important either by providing electrostatic or any other interaction.

Based on these experimental studies the short range interactions were investigated in
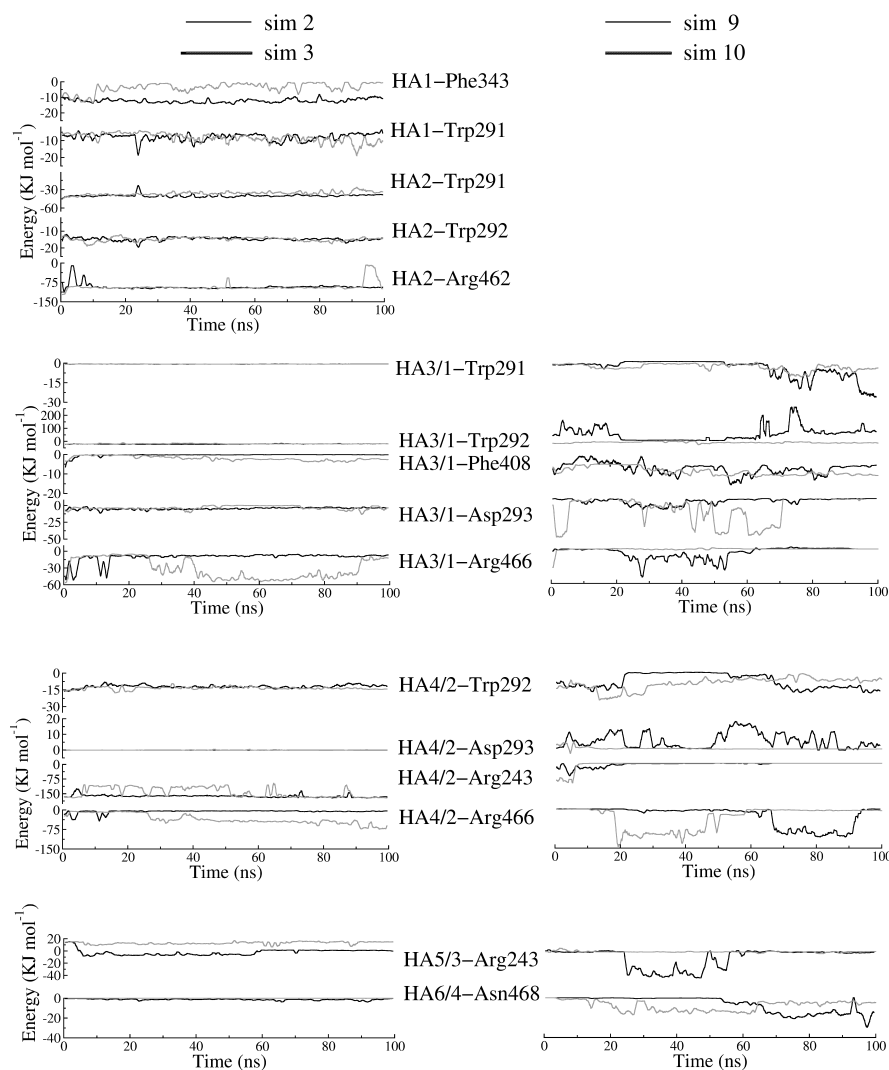
Figure 5.9: Short range interaction energies for the interactions of the important residues in the catalytic cleft of the *Spn*Hyal with the individual sugar rings. The labelling of the sugar ring HA3/1 means that the third sugar ring from the *holo* simulations (sim. no. 2 & 3) and the first sugar ring from the HLT3water simulations (sim. no. 9 & 10) is considered for the interactions, since these rings are in the similar positions inside the cleft. The sugar is stabilised inside the cleft by the balanced interplay of interactions with the residues positioned above and below it. Based on these interactions the roles of the known important residues can be validated as well as some new mutations can be proposed. For example, residue Trp292 contributing to the aromatic patch is known from the mutational experiments to play a key role and here in the simulations it is seen to provide different interactions to the third ring of the sugar depending upon whether the sugar is in 1loh-like position or in the processive phase. Similarly, residues Arg466 and Asn468 are seen to provide the attractive interactions for the sim. no. 9 and 10 and can be good candidates for the mutation experiments.

the four simulations discussed above (sim. no. 2, 3, 9 and 10). These include electrostatic interactions with each individual sugar ring 1 to 6 with specific individual residues particularly those residues for which mutation experiments were previously performed (Arg243, Asn349, Asn399, Tyr/Phe408 Trp291/Trp292/Phe343), and other specific residues of the cleft in contact with the sugar (from Fig. 5.8). Correlations between these interactions and the observations from Sec.5.3.6 were used to narrow down the number of residues significantly contributing to the interaction with the sugar. We performed this analysis with only the first 6 rings of *holo* simulations and the corresponding similarly positioned first four rings of HLT3water simulations, and not for the last two rings of HLT3water simulations (sim. no. 9 and 10). Figure 5.9 shows the significantly contributing ($\geq 10\text{kJ/mol}^{-1}$ energy) interactions of these interactions. As can be seen the balanced interplay of the interactions of residues positioned above and below the sugar keep the sugar stable inside the cleft. Of particular interest are the residues from the aromatic patch (Trp291, Trp291, Phe343), Asp293, and Arg243, which are positioned above the sugar, while Tyr408, Arg462, Arg466 and Arg468 are positioned below the sugar plane (See Fig. 5.8).

The residues in the aromatic patch (Trp291, Trp292 and Phe343) indeed play a vital role in the stabilisation of the sugar inside the cleft when it is in the catalytic phase. The interactions of these residues with the first two rings (HA1 and HA2) of sim.no. 2 and 3 are always stably attractive. For the third rings (HA3) of sim. no. 2 and 3 these interactions rapidly decrease to zero, while for sim. no. 9 and 10 the patterns are opposite. The interactions of Trp292 with the first ring of the sugar (positioned similar to HA3 of *holo* simulations) are of repulsive nature while the interactions with Trp291 provide attractive interactions. These interactions are in favour of processivity. Interestingly, however, Asp293 and Arg466 provide attractive interaction and may explain to some degree the lack of spontaneous processivity in these simulations.

Similar patterns are observed for the fourth ring of the sugar in *holo* simulations (HA4). Here, Trp292, Arg243, and Arg466 provide attractive interactions for the sugar in the catalysis phase. These residues interact differently with the similarly positioned second ring in HLT3water simulations (sim. no. 9 and 10). In this case, Trp292 provides weaker attraction while the interaction with Arg243 rapidly reaches zero. Taking these interactions into account Trp293 is suggested to provide the repulsive part of the interaction.

With these observations it seems that in addition to the residues that have been already verified experimentally to be critical in the binding/cleaving process (these include Arg243, Asn349, Asn399, Tyr408, Trp291, Trp292, and Phe343), also Arg466 and Arg462 may play a vital role in the processivity mechanism in general. Asn468 may also be a good candidate for mutational studies of the protein with longer substrates because this seems to provide the attractive interaction to HA6 (ring 6 of sugar in *holo* simulations).
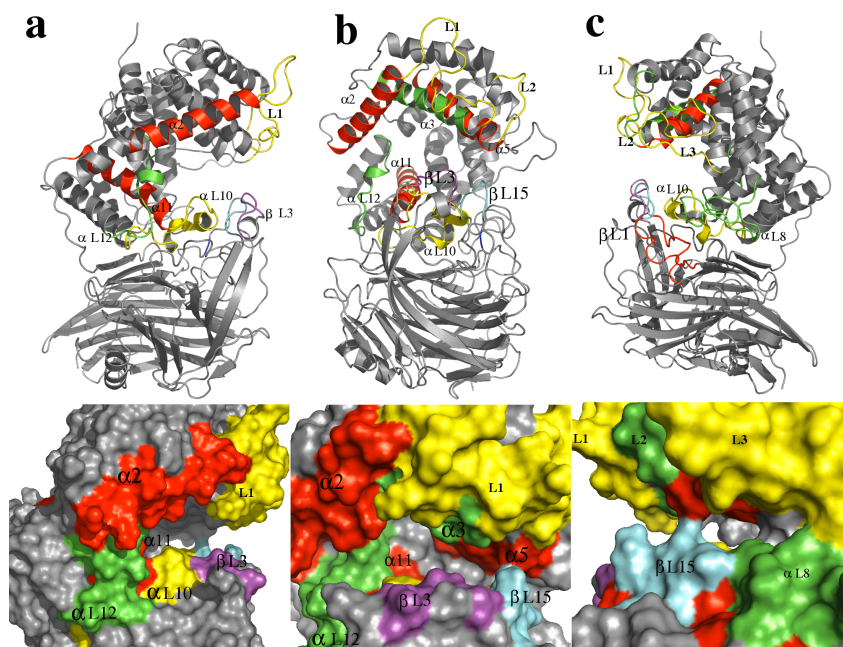
Figure 5.10: General regions around the catalytic cleft in the cartoon and the surface representation. (a) Entry-side region of the catalytic cleft consisting of 2 helices $\alpha2$ and $\alpha11$ (in red); loops in the $\alpha-$domain namely $\alpha L10$ (in yellow), $\alpha L12$ (in green) and part of loop L1 (in yellow); and a loop from the $\beta-$domain $\beta L3$ contributing to the cleft (in purple). (b) core-region of the catalytic cleft consisting of 2 helices $\alpha3$ (in green) and $\alpha5$ (in red), loops in the $\alpha-$domain namely L1 and L2 (in yellow), and a loop from the $\beta-$domain $\beta L15$ (in cyan). (c) Exit-side region of the catalytic cleft consisting of loops in the $\alpha-$domain namely L3 (in yellow) and $\alpha L8$ (in green), and a loop from the $\beta-$domain $\beta L1$ (in red) of which $\beta L15$ is part of.

5.3.8 Other Interactions

For the simulation 9 and 10, it was observed that the last two rings of the sugar (at the non-reducing end) move out of the plane away from an $\alpha-$helix at the entry side of the cleft of the protein (Fig. 5.6 C). This interaction might be an artifact of modelling the last disaccharide part into incorrect position. It would thus be interesting to monitor the interaction of the last disaccharide around this region since it may also provide necessary interactions for the longer substrates. To investigate this aspect and to characterise the interaction of sugar with the regions of the cleft at various stages of the processivity mechanism, we classify these regions as the entry-region, core-region and the exit-region of the cleft. These regions are shown in Fig. 5.10. The entry region contains residues from the positive patch, the core-region corresponds to the positive and the aromatic patch whereas the exit region contains the residues from the negative patch defined earlier. We are interested in investigating the interaction of the last parts of the longer substrates in the simulations (sim. no. 8 through 13) with the entry region of the cleft (Fig. 5.11).
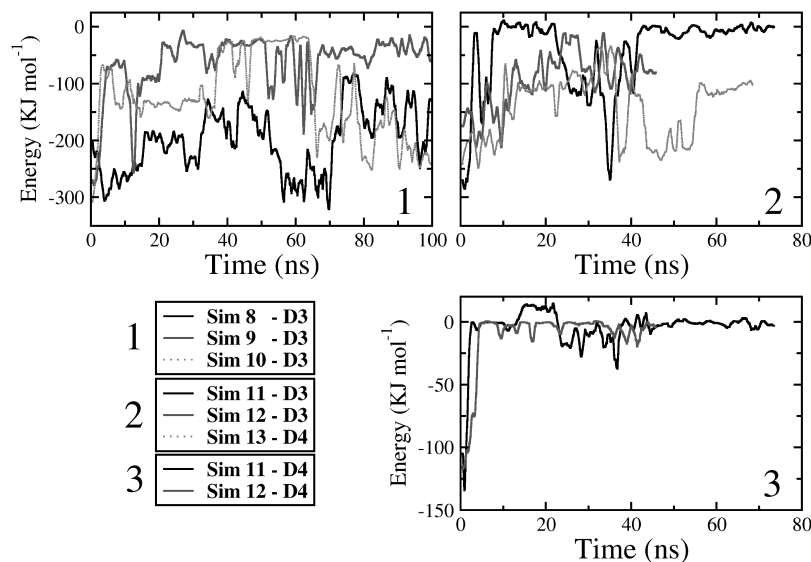
Figure 5.11: Short range interactions of the last disaccharide units (at the non-reducing end) of the HA with the entry region of the cleft. Upper figures (1 and 2) show the interactions of the last disaccharide for sim. no. 8—10 (D3) and 13 (D4). These interactions are initially very strong and then equilibrate showing that the HA has positioned itself in a favourable position after the manual modelling. The interactions are still attractive, but to varying degree. The picture at the right below shows the interactions of the last disaccharide unit (D4) of the sim. no. 11—12 with the entry region. Since this disaccharide is lying outside the cleft, the interactions decrease to zero quickly as expected.

On the top left in Fig. 5.11 the interactions with the last disaccharide unit (D3) at the non-reducing end of the HA are shown for the sim. no. 8–10. These interactions with the entry region of the cleft start with high values. Upper figures (1 and 2) show the interactions of the last disaccharide for sim. no. 8—10 (D3) and 13 (D4). These interactions start from very high values, probably because of the modelling of these units in unfavourable positions, and then decrease significantly showing that the HA has positioned itself in a favourable position. The interactions are still attractive as they should be, but to a varying degree. Similar behaviour is obtained for the sim. no. 11—13 where the substrate is one disaccharide longer. The picture on the upper right shows the interactions of the entry region with the penultimate disaccharide unit (D3) for sim. no. 11 and 12 and the last disaccharide unit (D4) for the sim. no. 13 (since the substrate is in 1loh-like position for the sim. no.13). The overall attractive interactions of all these units with the entry region of the cleft further validate the modelling of the last part of the HA in the long substrate. The picture at the lower right shows the interactions of the last disaccharide unit (D4) of the sim. no. 11—12 with the entry region. Since this disaccharide unit is lying outside the cleft, the interactions quickly decrease to zero as expected. The slight attractive interaction may be the contribution of the sugar ring7 in these cases, which is outside the cleft but near to the entry-region.

5.3.9   Complementary forces and origin of an energy barrier

With all the observations from the previous sections, we now can address the question of what keeps the sugar from sliding through the cleft in the simulations discussed in Sec. 5.3.4 or what is the binding force for the sugar when it is inside the cleft so that it does not leave the cleft even when the cleft is 'open'. As addressed in Sec. 5.1, these observations indeed point to two classes of interactions of HA inside the cleft in the processive phase (sim. no. 9 and 10) — repulsive interactions with the aromatic patch and weak attractive interactions with the positive patch. Even though the 1loh-like position is favourable for the sugar inside the cleft, spontaneous processivity was not observed in HLT3water simulations (sim. no. 9 and 10). This indicated the presence of an energy barrier for the processivity. The investigation of specific interactions such as electrostatic interactions, and hydrogen bonding with specific residues inside the cleft investigated for the two classes of the simulations also supported this view. One possible contribution to the interactions discussed in Sec. 5.3.6 may come from the hydrogen bonding of the sugar inside the cleft as discussed in 5.3.8. If that is the case, then it may also lead to the contribution to an energy barrier because for the probable reorientation of the sugar, these hydrogen bonds have to break. Thus, these hydrogen bonds of the sugar with the cleft also provide two opposite scenarios — keeping the sugar attached to the enzyme which is required for the processive nature of the enzyme, but at the same time presenting an energy barrier for the rotation phase of the sugar for proceeding to the next cycle.

The processivity of the enzyme may occur as an outcome of the dynamic competition between several interactions (electrostatic, hydrogen bonding or any other interaction), which are essential for providing strong binding forces so that the substrate does not leave the cleft and at the same time presenting an energy barrier for the process.

5.4   Conclusions

We have carried out various 'free' simulations of the *Spn*Hyal-HA complex in which the position of the HA substrate inside the cleft and/or its length is varied. From these extensive simulation studies the results from the previous chapter are confirmed. This gives sound support to the fact that bacterial Hyals are of flexible nature, which is very important in their physiological actions. In fact the conformational variance of the protein or the structural framework it provides is functionally important for the processivity and not just the random dynamical exploration of the energy landscape. The correlation of these interactions and the sugar movement inside the cleft with the domain motions of the protein indicate that the dynamics of the protein is actively involved in the processive mechanism. The role of the twisting mode is proposed to be responsible for the sliding of the sugar. This analysis on one hand provided a sound basis for the simulation model, while on the other hand provided some insights on the substrate affinity inside the cleft.

Investigation of the mobility of the sugar inside the cleft showed that even for the $\sim 100\,\text{ns}$ long simulations, a complete processivity by one disaccharide unit was not observed. This indicates that the process of sliding occurring after the catalysis is the main bottleneck or the time-limiting factor in the overall kinetics of the enzyme-ligand interaction. More detailed investigations are required to analyse the exact cause of this rate-limiting step.

It is seen that the substrate binding affinity, varies depending on the substrate position inside the cleft and also on the protein conformation. Roles of positive patch and aromatic of patch of the cleft are important in providing strong or weak attraction depending upon the substrate position. As the first step to understand the mechanism of processivity or sliding of the sugar through the cleft, the simulations were rationalised based on the available knowledge of mutation experiments, and the specific interactions between some of the known important residues and the HA-substrate were investigated in detail. These interactions have provided clues of how the substrate interacts when it is ready for the catalytic phase and when it is ready for the sliding phase. We have found, as expected, that the Trp291, Trp292, Phe343 (contributing to the aromatic patch of the cleft), Arg243, and Arg462 keep the first disaccharide of the sugar 'locked up' when it is in the catalytic phase by providing strong attractive interactions. For the rest of the substrate, varying interactions were observed depending upon the the position/length of the substrate and protein (whether it is open or closed initially). Here, also, the residues from the aromatic patch are observed to play a critical role in providing attractive interactions to the sugar helping it to remain stable inside the cleft. Also in this respect interactions of Tyr/Phe408 (of the catalytic cleft), Arg466 are seen to be important.

Analysis of the hydrogen bonding for all the simulations provided an additional contribution to the varying interactions of the sugar in the cleft. Based on the residue-level electrostatic interactions and hydrogen bonding analysis two more residues are observed to be important in the mechanism. Asn468 of the positive patch is seen to provide a strong attractive potential only for the longer substrates thus leading to a strong binding to the cleft. Similarly, Trp292 of the aromatic patch is seen to be involved extensively in the hydrogen bonding for all the simulations. The analysis of the role of these two residues provides better understanding of the binding mechanism of the sugar inside the cleft.

In conclusion, it can be said that the specific interactions and/or fluctuations in the protein-ligand complex and their correlation with the substrate movement show that the aromatic and the positive patch of the catalytic cleft are important in binding the substrate in between the multiple rounds of the catalysis. These patches are conserved in all the bacterial Hyals indicating their functional importance.

$$\langle e^{-\beta W}\rangle = e^{-\beta \Delta G}$$

T₂
T₁
T₀
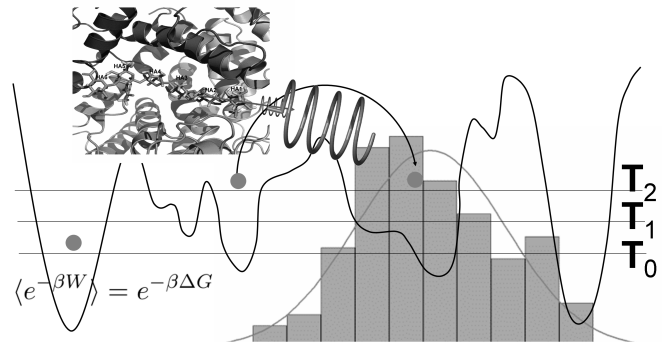
Chapter 6

Enforced MD simulations on *Spn*Hyals.

Summary

With the current state of the art computers, molecular dynamics simulations can probe the nanosecond to microsecond timescale dynamics of biomolecules. However, some of the processes, such as protein unfolding/folding or the sugar-translocation addressed in the present study, occur more slowly and the MD simulations need to be accelerated to span those regions of the energy landscape that the 'standard' MD simulation cannot access. Many different techniques of enhanced MD simulations are routinely used to drive the system towards a specific direction to induce the transitions of interest. These techniques differ in the algorithm, the method with which a bias force is applied, and also in the choice of the reaction coordinate along which the process is enforced to proceed. In the present work, we use primarily three such enhanced-MD techniques: High temperature simulations (Sec. 6.2), Force Probe Molecular Dynamics (FPMD) (Sec. 6.5) and Essential Dynamics (ED) (Sec. 6.6). Simulations from both these techniques concur that the cleft opening/closing motion and twisting of the domains of the protein relative to each other are essential for the processivity. Finally, we also present the preliminary free energy profile obtained for the enforced sugar translocation in Sec. 6.7.

6.1   Kinetic traps and force-induced simulations

Because of the complex structures of biomolecules, their energy landscapes are high-dimensional and of complex nature. The structure and dynamics of biomolecules strongly depend on these energy landscapes. Different motions/processes of the proteins can be described in terms of exploring the energy landscapes as shown in Fig. 4.1. Conformationally stable substrates of the protein are then represented by a large number of local minima separated by energy barriers of different heights. Local motions occurring on picosecond to few nanosecond timescales usually involve traversing of small energy barriers that may easily be studied with the modern day computation power. In fact the classical MD method has been very useful to investigate the properties of a system at multinanosecond timescales these days. However, many times the system under study gets trapped in a local minimum of the complex high dimensional energy landscape and crossing that barrier with the normal MD method usually requires a huge (and hence impractical) amount of computational time. A short MD simulation of a few nanoseconds, generally does not explore all those regions in the configurational space that are accessible under physiological conditions, where timescales usually are in the order of magnitudes larger. Sometimes, the rate-limiting factor of the complete process to be studied is on a relatively long timescale, requiring also very long, and hence impractical, simulation time. As a consequence, the conformational ensemble obtained by MD simulation is almost never complete. This is called 'sampling problem' and has been a general challenge for MD simulations for a long time. Several MD simulation methods to enhance conformational sampling have been developed during the last years.[45, 256]

In the present work, we do not observe any processivity from the 'free' (unbiased) simulations (Chapter 5). Hence, in this chapter we use three sampling-enhancement techniques to drive the system to the desired state: elevated temperature simulations, Force Probe Molecular Dynamics (FPMD), and Essential Dynamics sampling (ED). FPMD and ED sampling methods yield the forces on the system as a measure for the mechanical stability of the system. In general, they give insight into the properties of the system under mechanical load, *i.e.* under non-equilibrium conditions. In the following sections we discuss first the theoretical principles and then simulations using these methods in detail (Sec. 6.2 to 6.6). The particular questions that are addressed in the following sections are: (1) what are the critical pathways for the sugar translocation process? (2) what regions inside the cleft and/or which specific interactions of the cleft-residues with the sugar are responsible for the huge barrier that is not overcome in the long 'real time' simulations? and (3) how does the enzyme accomplish the directionality in relation to the characterisation of the earlier two questions? In other words, does the sugar overcome this barrier independently or is protein dynamics actively involved in the barrier modification, i.e. does the enzyme help the sugar to overcome the barrier?

6.2   Elevated temperature simulations

6.2.1   Principle
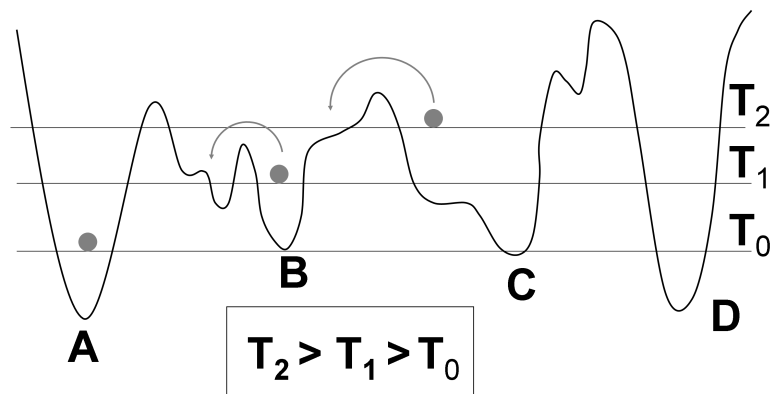
Figure 6.1: Schematic representation of the energy landscapes explored by the system (represented by grey sphere) at different temperatures. As the temperature of the system increases it becomes easier for the system to cross the energy barriers and go from one local minimum to another (say from C to B). The transitions that otherwise would not be possible, can thus be explored by increasing the system temperature.

Increasing the temperature increases the diffusivity. Therefore, one of the simplest methods to reduce the required simulation time, and thus to enhance the sampling efficiency, is to increase the temperature of the system of interest. This method has been successfully applied in computational studies of docking of ligands into the binding pocket of a number of proteins.[257] It has the advantage that the diffusion speed of the ligand is enhanced, thereby drastically reducing the simulation time required for searching appropriate binding modes, without affecting the dynamics of the proteins. The principle behind this technique is schematically depicted in Fig. 6.1. The system shown as a grey sphere at minimum A is at temperature T_0 and cannot explore any other part of the energy landscape because of the large energy barrier. A similar situation could occur if the system is at minimum B or C at temperature T_0. However if the temperature of the system is increased to T_1 or T_2 then the system can easily overcome the energy barriers as shown and explore the regions in the landscape that would not be accessible otherwise. Hence, we employ elevated-temperature simulations to study the change in the fluctuations of the HA substrate in the cleft as discussed in Sec. 6.2.2—6.2.4.

In the 'free' MD simulations employed in the present work the system was coupled to a heat bath by correcting the actual temperature using the *Berendsen thermostat*[141] (see Sec. 1.2.4). The coupling strength is set by the coupling constant τ, which was set to 0.1 ps. Temperature coupling assumes the system to be thermally equilibrated, and

removes temperature differences. In contrast, in the elevated temperature simulations part of the system has a different temperature. To achieve this, the coupling constant τ was decreased to 0.01 ps for the simulations presented here.

6.2.2 Simulation setup

Four different simulations were setup to test the high-temperature behaviour of the sugar inside the cleft of *Spn*Hyal. These simulations were carried out with the MD software package GROMACS 3.2.1 and GROMACS 3.3.[82,83] The starting structure for each simulation was the same as that of the 'free' simulation corresponding to that system. As discussed in Sec. 6.2.1, the temperature coupling constant τ was decreased to 0.01 ps. The remaining details of these simulation systems were similar to those set for the 'free' simulations as discussed in Sec. 4.3. The details of the simulation systems is sketched in Table 6.1. The simulations correspond to the free simulation configuration 2,4 and 10 of Chapter 5. We chose these systems as in these systems the protein showed the least opening of the cleft (see Sec. 5.3.2 and Fig. 5.2). The only difference to those simulations is the now increased temperature of the sugar as indicated in the Table. 6.1. In all four simulations the protein and solvent are kept at 300 K. The equilibration time for simulations was now increased to 10 ns for the system to acclimatise for the changes in the substrate structure.

No	Simulation	substrate length	Temperature of the sugar (K)	time (ns)	Total System size (number of atoms)
1.	*Spn*Hyal-holo	hexasaccharide	500	100 (10)	83,542 (10)
2.	HLT2water	tetrasaccharide	500	100 (10)	89,550 (9)
3.	HLT3water1	hexasaccharide	500	20 (10)	100,486 (10)
4.	HLT3water2	hexasaccharide	750	36.5 (10)	100,486 (10)

Table 6.1: List of high-temperature simulation systems simulated. The simulation time T in nanoseconds is stated with the equilibration time in the brackets. Similarly, in the column of total number of atoms N in the system the number in the brackets gives the number of counterions (Na^+) used to neutralise the system charge.

6.2.3 Results and discussions

Two main results were compared to those in the case of room-temperature simulations (Chapter 4 and 5). First, the PCA projections of the protein were compared to see how the protein responds to the sugar at high-temperature (see Fig. 6.2). Panels A and B show the 2-dimensional projections of the simulation trajectories on the first two eigenvectors as obtained by the PCA of the crystallographic data (on the left), and on the right is the similar projection on 2nd and 3rd eigenvector. Panel C shows the
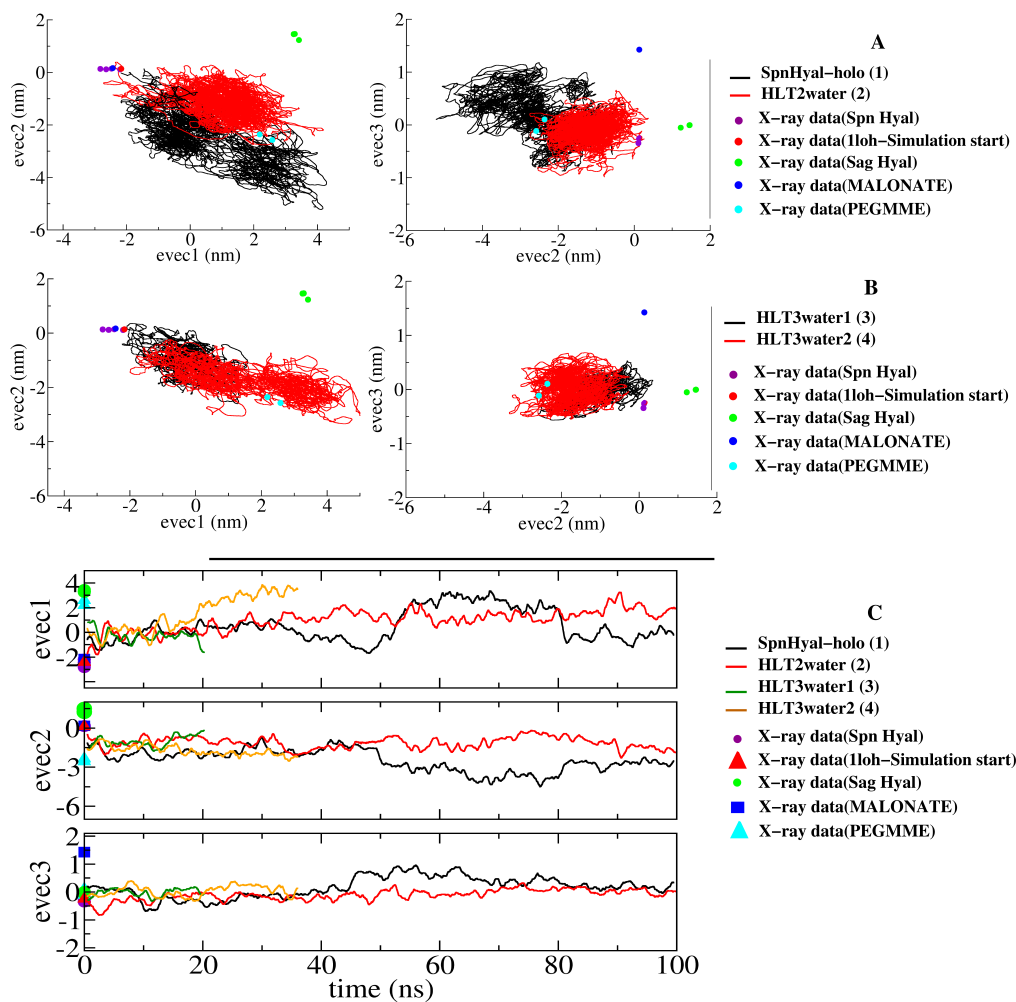
Figure 6.2: A,B: 2-D Projections of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. On the left hand side the projections along the first two eigen-modes show the extent of the opening/closing of the cleft (along the x-axis) and the twisting of the α−domain with respect to the β−domain. The opening/closing of the entrance/exit of the cleft access (projection along the third eigenvector) is shown along the y-axis on the right. The flexibility of the protein increases significantly from the respective cases of the simulations in Chapter 5. The *holo*-simulation (black trace in A) even showed the opening of the cleft which was not observed during the 'free' *holo* simulation. These plots indicate a strong coupling of the protein dynamics with the sugar mobility inside the cleft. C: Time evolution of the simulation trajectories together with the x-ray structures projected on the individual eigenvectors obtained from the PCA. The x-ray structures of Hyal are shown as circles and the simulation trajectories are shown as traces.

time evolution of the simulation along the individual eigenvector. Table 6.2 gives the quantification of these eigenvectors in terms of amplitudes of cleft opening or twisting motion or cleft/access opening similar to Table 5.2.

Sim. No.	Eigenvector (i) (opening/closing of the cleft)	Eigenvector (ii) (twisting of the $\alpha-$ domain)	Eigenvector (iii) (opening/closing of the access to the cleft)	Eigenvector (iii) (opening/closing of the exit to the cleft)
1	12.7	6.6	5.3	6.2
2	9.4	4.3	3.4	11.2
3	3.9	2.9	3.0	4.0
4	9.8	4.4	3.3	5.5

Table 6.2: The extent of the dynamic domain motions in all the simulations (same as described in Chapter 4). The order of the simulations is the same as shown in Table 5.1. For the first eigenvector the distance between the $C^{\alpha}-C^{\alpha}$ separations of Asn231 and Gly769 in the most open and least open trajectory snapshots gives the extent of the opening/closing motion of the cleft. The twisting extent (amplitude of the second eigenvector) is measured by the separation between the C^{α} atoms of Asp340 for the superimposed most twisted and the least twisted snapshots from the trajectory. The last two columns give the extent of the amplitudes for the third eigenvector showing entry-side opening and exit-side opening respectively. For the entry site opening the distance between the $C^{\alpha}-C^{\alpha}$ separations of Asp211 and Ser771 is measured while for the exit side opening the distance between the $C^{\alpha}-C^{\alpha}$ separations of Asn341 and Asn580 is measured to give the extent of the opening/closing of the access to the cleft.

Interestingly, from these observations, it is convincingly seen that the protein domain motions are affected by increasing the temperature of the sugar in each case. In each simulation the protein opened to larger extent than in simulations with the sugar at 300 K in the respective case (See Fig. 5.2, Fig. 5.3 and Table 5.2). The amplitudes for all the eigenvectors in sim. no. 3 are smaller than that obtained for the simulations at 300 K, but that can be attributed to the fact that this simulation was carried out for only 20 ns. Surprisingly, only the first eigenvector seems to be affected in all the simulations. The amplitudes of the motions for the second eigenvector i.e. the twisting motion was observed to be almost of the same extent as observed in Chapter 5. The opening/closing of the access to the cleft, i.e. the opening of the entry region of the cleft was observed to be to a lesser extent than before.

Next, the flexibility of the sugar inside the cleft and its movement in the processive direction were monitored to see whether a complete translocation of the sugar is observed during these simulations. As can be seen from Fig. 6.3 (a), the root mean square deviation (RMSD) of the sugar clearly increased compared to that at 300 K, showing that the sugar at high temperature is more dynamic, as expected. The RMSD was obtained in a similar manner as in Chapter 5 (see Sec. 5.3.3). This gave a measure for the flexibility of the first four rings of the sugar inside the cleft. In all simulations the overall flexibility of the sugar was increased due to the higher temperature, but it remained stable fluctuating at these high values. The simulation with the hexasaccharide substrate at 750 K (sim. no. 4) showed a decrease in the RMSD towards the 1loh-like sugar initially but then remained stable. The sugar in simulation 1, i.e. with the sugar in the catalytic position showed a very dynamic character. The RMSD was correlated
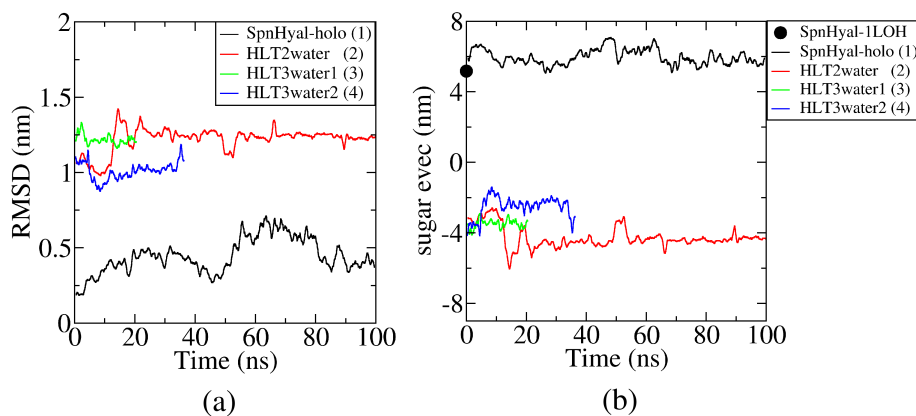
Figure 6.3: (a) Flexibility of the first four rings of HA inside the cleft. RMSDs are calculated (in similar manner as in Chapter 5) with respect to the first four rings of HA in the crystal structure 1loh, which is used as the starting structure for all the simulations in this work. In all the simulations the overall flexibility of the sugar was increased due to higher temperature, but it remained stable (except in *Spn*Hyal-*holo* case). The simulation with the hexasaccharide substrate at 750 K showed a decrease in the RMSD towards the 1loh-like sugar initially but then remains stable. The sugar in simulation 1, i.e. with the sugar in the catalytic postion shows a very dynamic character. This results in the opening of the cleft, which was not observed in the 'free' simulations before. (b) Motion of the sugar inside the cleft obtained by the PCA performed on the sugar in a similar manner as in Chapter 5 (see Sec. 5.3.4). The black circle shows the position of the sugar in the *Spn*Hyal structure. For one complete processive cycle, the simulations ready for the sliding phase (sim. no. 2 to 4) are expected to show motion towards this circle (by moving through the distance of one disaccharide unit and attaining the 1loh-like structure). The black trace shows the simulation trajectory of *Spn*Hyal-*holo*. The sugar in this simulation is stable as expected. None of the simulations shows complete translocation towards the processive direction, supporting the proposed energy barrier in the processive mechanism.

to the opening of the cleft, which was not observed in the 'free' simulations before. The PCA on the sugar was carried out to see whether any movement of the sugar in the processive direction was observed. As can be seen from Fig. 6.3 (b) there is no spontaneous processivity observed during these simulations. Hence, even though the flexibility of the sugar is increased, there is still some barrier that keeps the sugar 'locked' up in its positions.

6.2.4 Conclusions

Results indicate that modification in the behaviour of the sugar is inherently coupled to the dynamics of the protein. This supports the earlier observations of the dynamic interplay of the protein and the sugar in the processivity mechanism and reconfirms the finding that the protein has to accommodate/modulate its function according to the conditions of the sugar. This would also support the hypothesis that protein domain

motions are functionally important for the sugar translocation/processivity. The increased flexibility of the sugar correlates well with the opening of the cleft of the protein especially for the *Spn*Hyal-*holo* simulation, in which the protein remained 'closed' in the free simulations. However, even though the dynamics of the sugar increases it is not sufficient to drive it towards the catalytic position if it is positioned one disaccharide unit before. This may support the proposed energy barrier from Chapter 5. Since the sugar is still not able to cross this barrier, we now need more sophisticated procedures to drive the system forcefully in processive direction to study the mechanism in detail.

6.3 Force Probe Molecular Dynamics (FPMD) simulations: Principle

In high temperature simulations, as described above, the system freely explores the energy landscape at that temperature. Furthermore, the elevated temperature may lead to artefacts due to temperature coupling, particularly because in our case the temperature of only the subsystem (HA) was enhanced. The larger the difference between the temperature of the HA and the *Spn*Hyal, the larger are the artefacts expected to be. Therefore, it is not desirable to raise the temperature to a very large extent. As could be seen from Sec.6.2, we obtained evidence that flexibility or mobility of the sugar is inherently coupled with the protein dynamics. However, the complete processivity through one disaccharide unit was still not obtained by this increased but 'free' diffusion. To this end, it is necessary however, to drive the system in a particular direction to effect the desired transition. In FPMD simulations the system can be forced to the desired state by 'pulling' one or more pull groups or parts of the system in a pre-defined pulling direction specified in cartesian coordinates. This mimics the situation in an AFM experiment (for more details on the AFM technique see Appendix C) by employing a harmonic potential (or a 'virtual' spring) that acts as a flexible cantilever. This technique allows the system to explore the areas of the configurational space it otherwise would not visit via 'real-time' MD simulation. To account for statistical errors, series of AFM experiments are performed to yield only mean values of the results (say forces of individual rupture events). Comparison to molecular simulations is therefore straightforward. Thermodynamic and kinetic quantities, such as free energies and activation free energies of folding and binding, obtained from the molecular simulations can in principle be directly compared to the experiments, provided these simulations allow the estimation of these equilibrium properties.

Fig. 6.4 shows an example of a force-probe MD simulation system. Pulling is carried out by subjecting one or more atoms i of the simulation system to a harmonic spring potential,

$$V_{\text{spring},i}(t) = k_0[z_i(t) - z_{\text{spring},i}(t)]^2, \tag{6.1}$$

where k_0 is the spring constant, $z_i(t)$ the position of the center-of-mass (COM) of the pulled atoms i, and $z_{\text{spring},i}$ the position of the spring.
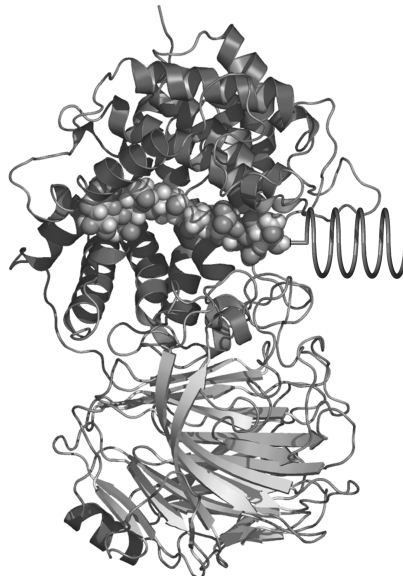
Figure 6.4: A typical FPMD simulation consists of an atom or group of atoms subjected to a harmonic spring potential. In this example, HA (sphere representation) is 'pulled' through the cleft of the *Spn*Hyal. The COM of the protein cleft is kept fixed at all the time.

As the spring moves with constant velocity v along the specified direction mechanical stress on the pulling group(s) is developed. The spring position can be calculated according to

$$z_{\text{spring},i}(t) = z_i(0) \pm \Delta z(t), \tag{6.2}$$

where $\Delta z(t) = vt$ is the spring dislocation. Forces on each atom of the pull groups as they move under the influence of the spring are

$$F_i = k_0 \left[z_i(t) - z_{\text{spring},i}(t) \right]. \tag{6.3}$$

The force profiles (force plotted against either time or spring position) give insight into the overall behaviour of the system under mechanical load similar to single molecule AFM experiments. In this thesis, we simulated the force-induced translocation of the HA-substrate (Sec. 6.5) by subjecting the first sugar ring at the reducing end of the HA-substrate to a harmonic potential while keeping the center-of-mass of the catalytic cleft fixed. Force profiles are obtained using Eq. 6.3 and compared for different conditions. Maxima of the force profiles denote rupture forces corresponding to the phenomena where the sugar starts moving along the chosen reaction coordinate.

FPMD simulations are becoming increasingly popular and find many applications.[258–260] One major disadvantage of this method is however, that the reaction coordinate along which the system is driven is defined externally but in most cases this coordinate is not known *a priori*. Choosing an appropriate reaction coordinate is both, very important

and a challenging task for such simulations and can only be partially solved by sampling many directions and comparing the corresponding force profiles. In the case of the *Spn*Hyal-HA system we chose the vector from the non-reducing end towards the reducing end of the sugar as the pulling direction.

We use the established technique of FPMD simulations to 'pull' the HA-substrate along an appropriate reaction coordinate in the (putative) catalytic/processivity direction. To investigate the influence of the protein domain motions observed in the previous chapters, we use these FPMD simulations independently as well as in conjunction with another sampling enhancement technique called Essential Dynamics sampling technique (EDSAM), which is described in the next section.

6.4 Essential Dynamics (ED) Sampling: Principle

In Essential Dynamics (ED) sampling, another efficiency enhancing sampling technique,[27, 28] the reaction coordinate can be deduced from the PCA (Sec. 1.3.1) of a short free MD simulation of the system under investigation. The reaction coordinate represents one or several of the principal modes of motion resulting from PCA of the system. Subsequently, the system is driven along this reaction coordinate. This way external parameters are minimised since now the reaction coordinate is also an inherent property of the system. In other words the system is enforced to undergo functional motions as deducted from its dynamics at equilibrium (i.e. from free dynamics)
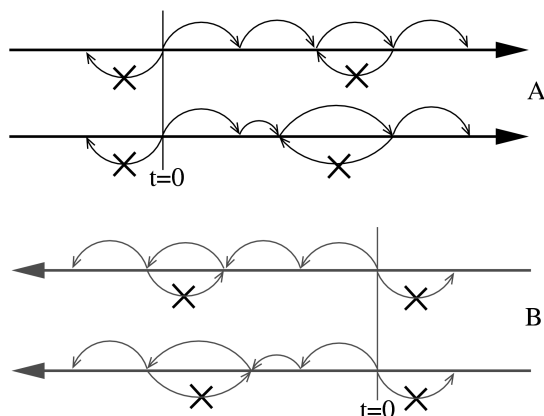
Figure 6.5: Schematic representation of Essential Dynamics technique for two linear sampling methods along a chosen eigenvector in forward (A) or backward (B) direction. In one method the stepsize of the increment is fixed (top in each case) while in the other method the stepsize is not fixed and free to vary (bottom in each case). Figure adopted and modified from.[261]

Figure 6.5 shows how ED sampling is put into practice. One or more eigenvectors are defined as the subspace in which sampling shall be enhanced. Then, one of two ED sampling variants can be applied: (1) by specifying the direction along the principal mode and the fixed stepsize towards the desired end-state, or (2) by specifying the

direction along the principal mode but no fixed step-size. In this case, all step-sizes were allowed. In both variants only those steps were accepted that were not along the negative direction of the reaction coordinate. The other degrees of freedom are unperturbed and can equilibrate during the ED simulation. In both variants the parameter that determines the simulation time for ED sampling is the step size. However, in the latter an optional upper limit for the progressive motion may be given by choosing an upper limit for the step size.

We have used this technique to force the HA-substrate to the target position independently as well as in conjunction with the FPMD simulations to study the role of protein motion with respect to sugar translocation. Both these techniques stimulate the system and the system is no longer in equilibrium conditions. The forces we obtain from these methods however, can be used to derive the equilibrium properties such as the free energy profile of the system, which describes the overall energy barriers to be overcome during the process.

6.5 FPMD simulations: Results and discussions

Using the established FPMD simulation technique (Sec. 6.3), we carried out a series of FPMD simulations with an additional external force applied to the HA substrate 'pulling' it through the binding cleft of *Spn*Hyal. During this process, the pulling force was monitored as a function of the travelled distance.

6.5.1 Rupture force, pulling speed, and force constant

All the simulations were performed using the GROMACS software package 3.2.1.[60, 82] The starting structure for all enforced calculations studied here is the one with the hexasaccharide substrate inside the cleft of the *Spn*Hyal in the position after catalysis (*i.e.* ready for the sliding, or similar to sim. no. 9 or 10 from Chapter 5). The basic simulation set up is similar to that discussed in Sec. 4.3. The harmonic spring was positioned at the first ring (HA1) of the sugar at the reducing end and the pulling directions were chosen along the direction of the vector connecting the COM of the first ring (HA1) and the COM of the second ring (HA2) (see Fig. 6.6). This pulling direction was chosen such that the sugar was pulled directly towards the catalytic site to acquire a 1loh-like structure. To achieve this the sugar has to move by one disaccharide unit, *i.e.* about 1 nm. To this end, all the simulations were carried out until the sugar had moved at least 1 nm in the forward direction. The simulation times ranged from 400 ps (for the stiffest spring $k = 5000$ kJ/mol/nm^2 and fastest velocity of 1 nm/100 ps) to 20 ns (for the simulations with a spring constant of $k = 250$ kJ/mol/nm^2 and the slowest velocity of 1 nm/5 ns) totalling more than 100 ns of FPMD simulation time in this set. The center of mass (COM) of the cleft of the protein was kept fixed at all times to avoid the rotation of the protein induced due to the applied pulling.
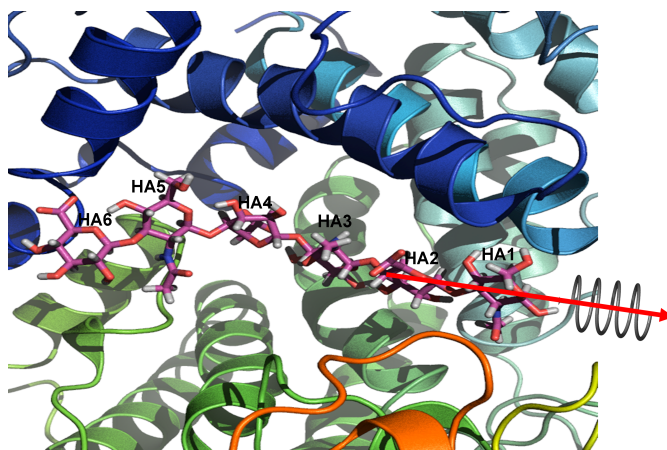
Figure 6.6: Closeup of the region around the cleft of the protein (cartoon representation) together with the sugar (stick representation). The arrow and the spring denote the pulling direction for the simulation to enforce the sugar translocation.

Initially, a number of simulations with different pulling velocities v and different force constant for the harmonic spring k were performed to 'pull' the sugar in the forward (processive) direction:

- $k = 100$ kJ/mol/nm^2 : 1 nm/1 ns (twice), 1 nm/2 ns, 1 nm/5 ns

- $k = 250$ kJ/mol/nm^2 : 1 nm/100 ps, 1 nm/1 ns (twice), 1 nm/2 ns,1 nm/5 ns

- $k = 500$ kJ/mol/nm^2 : 1 nm/100 ps, 1 nm/500 ps, 1 nm/1 ns, 1 nm/2 ns, 1 nm/5 ns (all simulations twice)

- $k = 750$ kJ/mol/nm^2 : 1 nm/100 ps, 1 nm/1 ns, 1 nm/2 ns (all simulations twice)

- $k = 1000$ kJ/mol/nm^2 : 1 nm/100 ps (twice), 1 nm/1 ns (3 times), 1 nm/2 ns

- $k = 5000$ kJ/mol/nm^2 : 1 nm/100 ps (twice), 1 nm/1 ns (3 times), 1 nm/2 ns (3 times), 1 nm/5 ns

Of these, $k = 1000$ kJ/mol/nm^2 and $k = 5000$ kJ/mol/nm^2 correspond to an almost infinite stiff spring behaviour (which may be good for the calculation of the potential of mean force (PMF) in conjunction with Jarzynski identity see Sec. 6.7.3), while $k = 100$ kJ/mol/nm^2 was very soft requiring a large amount of time to build up enough force on the sugar to actually move. In general, the slower the speed, the better is the agreement with the free simulations. However, to increase the efficiency the speed is increased appropriately and the force constant is chosen high enough (to make the spring stiff enough spring so that the pulling group follows it almost instantaneously). The pulling speeds mentioned above are still much larger than normally encountered

101

in experimental studies. However, the approach is still valid as long as it is made sure that we can extrapolate from the simulation regime to the experimental regime.[262] So far, there are no studies performed with the long HA being pulled through the cleft experimentally and therefore without the experimental feedback one cannot be sure of the pulling speeds employed or the values of the forces obtained from these force profiles. However, it is to be expected that experimental pulling speeds would be orders of magnitude smaller.

Here, with FPMD simulations, we wish to study the overall behaviour of the protein-sugar complex in response to mechanical forces, and whether there is any hint towards an energy barrier as indicated by the free simulations in Chapter 5. Choosing the speed or force constant is not a very critical issue as long as it does not impose a change of the pathway. This can be ensured by choosing the pulling velocities from the thermally activated regime (*i.e.* a logarithmic dependency of the pulling velocity on the rupture forces[262]). This was indeed seen in Fig. 6.7 (D) that shows the maximum rupture-forces for all the simulations mentioned above. The plot shows a linear dependence for high pulling velocities ($v > 1$ nm/ns), showing the friction-dominated pulling for these velocities, and a logarithmic dependency for slower velocities ($v < 1$ nm/ns), the thermally activated regime. Another criterion to chose the values of the pulling speed and the force constant in the present studies may be the comparison of the protein behaviour as a response function to these parameters. Since we have fairly good confidence in the functional role of protein domain motions from the sub microsecond simulations, we checked in each of the above simulations the protein domain motions as obtained from the projection of the simulation trajectory on to the PCA performed on the set of the experimental structures (same procedure as was carried out in Sec. 4.4.2). Comparing the domain motions along the first three eigenvectors in each of the simulation to those observed in the *Spn*Hyal-apo simulation (not shown), we chose the combination of $k = 750$ kJ/mol/nm^2 and $v = 1$ nm/ns for the stiffness of the spring and the pulling speed respectively. This way, we made sure that if protein conformational changes are indeed coupled to sugar translocation, the simulations are long enough to allow for such changes.

Figure 6.7 (A) shows a typical force curve obtained from the FPMD simulations. As the spring moves along the chosen coordinate, the force on the sugar starts building up and the rupture force gives the measure of the maximum force required for the sugar to actually start moving along the reaction coordinate. Fig. 6.7 (A) shows the force curves for one set of the simulations described above (for $k = 750$ kJ/mol/nm^2). For the force curve corresponding to the simulation with pulling speed $v = 1$ nm/ns, the sugar starts moving at 2 ns and the overlay of the initial structure and the structure at 2 ns shows complete movement of the sugar by one disaccharide unit (B). The projections of the simulations for this force constant on to the PCA eigenvector set and the comparison to the *Spn*Hyal-*apo* simulation carried out previously, show that a pulling speed of $v = 1$ nm/ns is fairly suitable as the pulling speed since it allows collective protein motions to relax as a response to the sugar translocation (C). We chose this set of spring constant and the pulling speed for our next simulations.
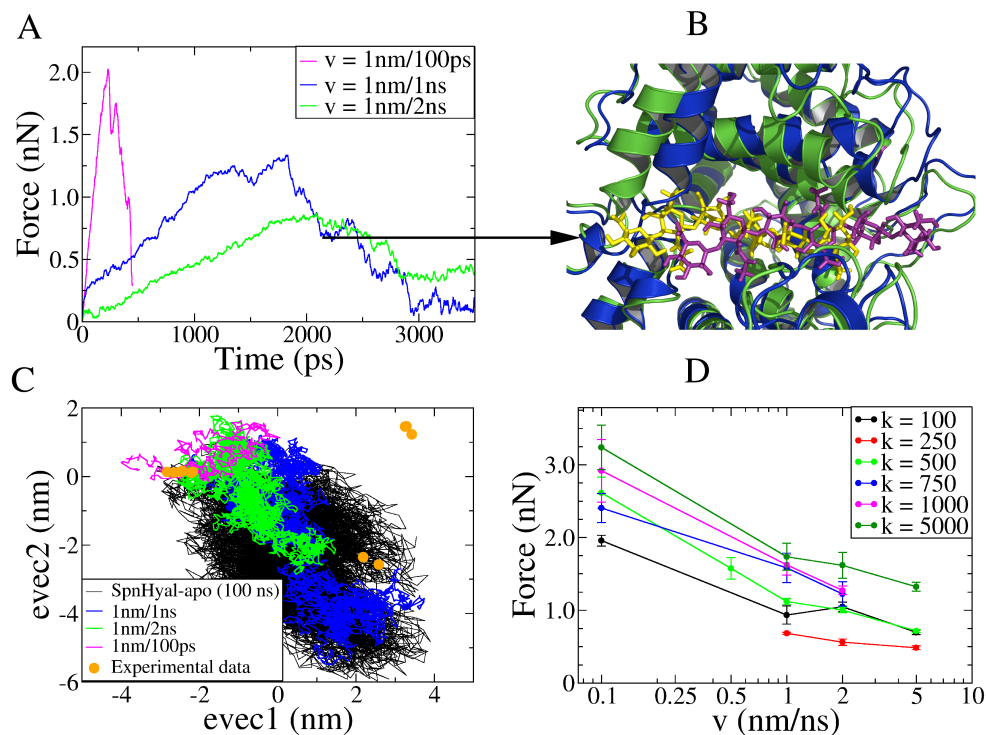
Figure 6.7: (A) Force curves for three different pulling speeds in the simulations with $k = 750$ kJ/mol/nm^2. The maximum force, called the rupture force, shows the event when the sugar moves through one complete disaccharide unit in the expected direction. This is shown in B, with the overlay picture of the initial structure (protein in green cartoon and the sugar in yellow stick representation) and the structure at 2 ns (protein in blue cartoon and the sugar in purple stick representation). In this picture the protein at 2 ns can be seen to be 'open' and 'twisted' showing the protein response to the sugar pulling. The projection of the simulation trajectories on to the first two eigenvectors obtained from the PCA of the experimental structural data is shown in panel C (see also Chapter 4). Here it is clearly seen that as the sugar is subjected to the FPMD simulation, the protein responds by opening and twisting. The magnitude and direction, however, depends on the pulling speed. By comparing the extent of the opening and twisting to those obtained from the SpnHyal-apo simulation, we chose the pulling speed of $v = 1$ nm/ns for the rest of the pulling simulations. Panel D shows the maximum rupture forces for all the simulations performed in this set.

6.5.2 Protein flexibility in FPMD simulations

To obtain a better sampling of the pulling simulations with the chosen parameters ($k = 750$ kJ/mol/nm^2 and $v = 1$ nm/ns), we performed FPMD simulations independently as well as in conjunction with the EDSAM technique, to pull the sugar in the forward as well as in the backward direction to address the question whether the protein conformation affects sugar translocation. To this end, two types of structures were chosen. Simulations for the first set were performed on the 'closed' structure of the protein

103

(1loh-like) and the sugar was then pulled under three conditions:

- Free exploration of the protein dynamics. (Forward),

- Fixed individual PCA eigenvector coordinates of the protein. Here, keeping the first eigenvector fixed resulted in the cleft 'closed' at all times (ED1), keeping the second eigenvector fixed did not allow the protein to twist (ED2), and keeping the third eigenvector fixed resulted in keeping the entry/exit of the cleft locked in the original position (ED3).

- Pulling the sugar backwards in the non-processive direction such that it emerges from the cleft at the non-reducing end (Opposite). The pulling direction for these simulation was chosen as the vector connecting the COM of the last ring (HA6) and the COM of the penultimate ring (HA5). These simulations were performed to obtain any knowledge on the directionality of the processive mechanism.

The principle of these simulations is sketched in Fig. 6.8. Similarly, all simulations described above were also performed with the 'open' structure of the protein. To this end, the 'open' structure configurations were first obtained by EDSAM simulations that drove the protein to the 'open' configuration (see Sec. 6.4) while keeping the sugar fixed as in the original configuration (such that it does not move during these simulations). Subsequently a similar set of FPMD simulations was performed on each of the 'open' conformations.

Simulations for the 'closed' structure were performed on ten different configurations, whereas simulations on five different 'open' configurations were performed. Figure 6.9 shows the result for representative cases for each 'closed' and 'open' scenario. Panel A and B show the force curves for the simulations in each scenario. They show the development of the forces on the sugar until it starts moving in the desired direction causing forces to drop suddenly to very low values. These rupture forces vary under different conditions depending upon whether the pulling is performed in forward or opposite direction. Panel C and D show the 1D projection of the FPMD simulation trajectories projected to the individual eigenvectors, along with the *Spn*Hyal-*apo* simulation carried out previously (see Chapter 5). These plots allow the comparison of the extent of protein domain motions in each case.

A major difference in the force profiles in the panels A and B is the overall lower profile for the simulations starting with 'open' structure compared to those with 'closed' structures. This shows that the opening of the protein cleft facilitates (and may be necessary for) the translocation of the sugar. This can be further validated by the projections of these simulations on the first eigenvector (panels C and D, evec1-plot). Here it can be seen that the protein opens up significantly in all the simulations starting with the 'closed' structure. For the simulations starting with the 'open' structures the protein was already open and remained open for the duration of the simulations. The opening in already open scenario was observed more pronounced than that observed in
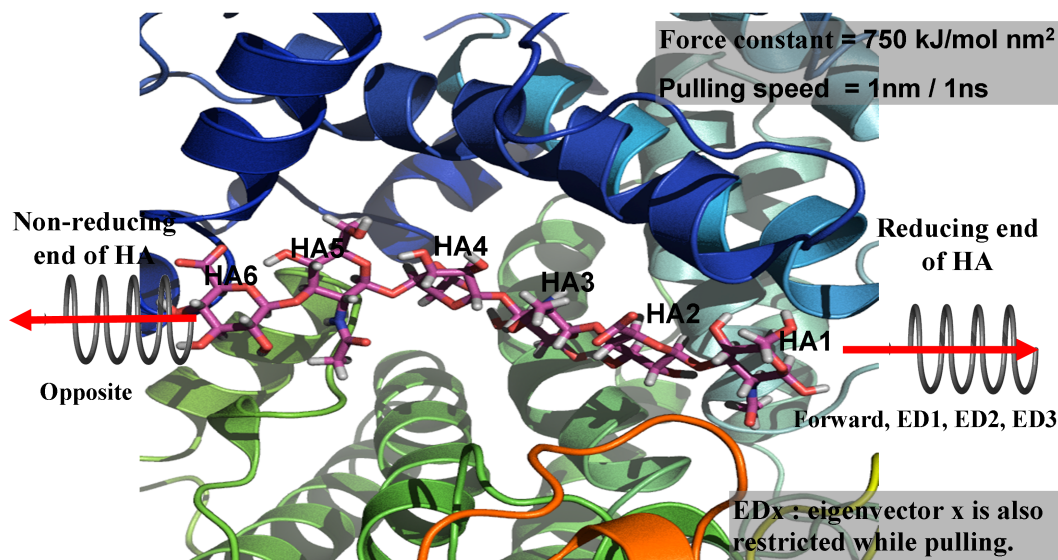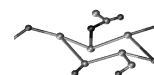
Figure 6.8: Scheme of the FPMD simulations setup described in Sec. 6.5.2. The arrow at the "non-reducing" end (left) of the sugar depicts the pulling of HA in the 'opposite' direction (in the non-processive direction). Similarly, the arrow at the reducing end of the HA (right) denotes the pulling in the putative processive direction for four cases (pulling with protein free and pulling with each of the eigenvectors of the protein fixed).

apo-simulation. Interestingly, in each scenario the simulations with the first eigenvector 'fixed' also concur with this observation (in both panels A and B, ED1 depicts the forces for the pulling simulations with fixed eigenvector 1, red curve). For the simulation with the protein cleft kept closed and fixed at all times (A, red curve) larger forces were required for the translocation of the sugar compared to all other cases. On the contrary the simulations performed by forcing the protein cleft 'open' at all times showed lower rupture force than the corresponding Forward pulling (B, black and red curves) thus strongly supporting the functional role of the opening mode.

While the relation between sugar mobility and protein conformation for the opening mode was clear and consistent in all the 15 simulations, the relation with the second eigen-mode was observed to be varying for the simulations in the 'closed' protein scenario. In panel A of Fig. 6.9 it is seen that forces required for the simulation where the twisting of the protein was not allowed (ED2, green curve), were lower than those observed when the protein is free (Forward, black curve). This, however, is may not an indication that for the translocation the twisting of the protein is not favourable, as in
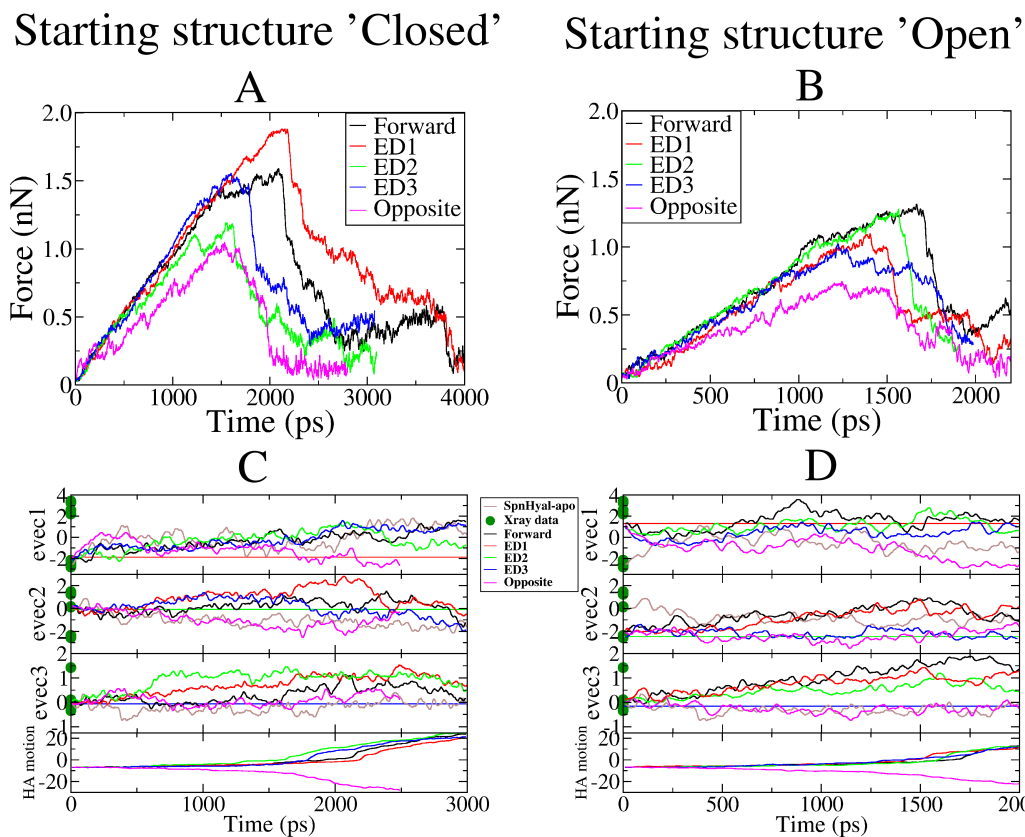
Figure 6.9: (A) Force curves for the pulling simulations with the starting structure of the protein 'closed': Forward — sugar pulling with protein free, ED1 — sugar pulling with the first eigenvector of the protein kept fixed at all times, ED2 — sugar pulling with the second eigenvector of the protein kept fixed at all times, ED3 — sugar pulling with the third eigenvector of the protein kept fixed at all times, Opposite — sugar pulling in the opposite (non-reducing) direction. (B) Similar force curves for the pulling simulations with the starting structure of the protein as 'open'. The overall lower force profiles and time-durations required for the translocation of the sugar indicate that opening of the cleft of the protein is favourable (and may be necessary) for the processivity. The lower forces for the opposite pulling concur with the presence of an energy barrier as proposed previously. (C) and (D): The projections of the simulations to the individual eigenvectors obtained by PCA versus time, as discussed in Chapter 4.

three of the other simulations in the 'closed' protein scenario (not shown), the rupture forces were larger than the "free-Forward" simulations (for others they were comparable). The variation of this motion in the pulling simulations may reflect the complex nature of the interplay of this domain motion with the first eigenvector (opening of the cleft) and the translocation of the sugar. To obtain a more detailed view on the functional role of this domain mode, more statistics is needed from additional pulling simulations (with possibly slower pulling speeds).
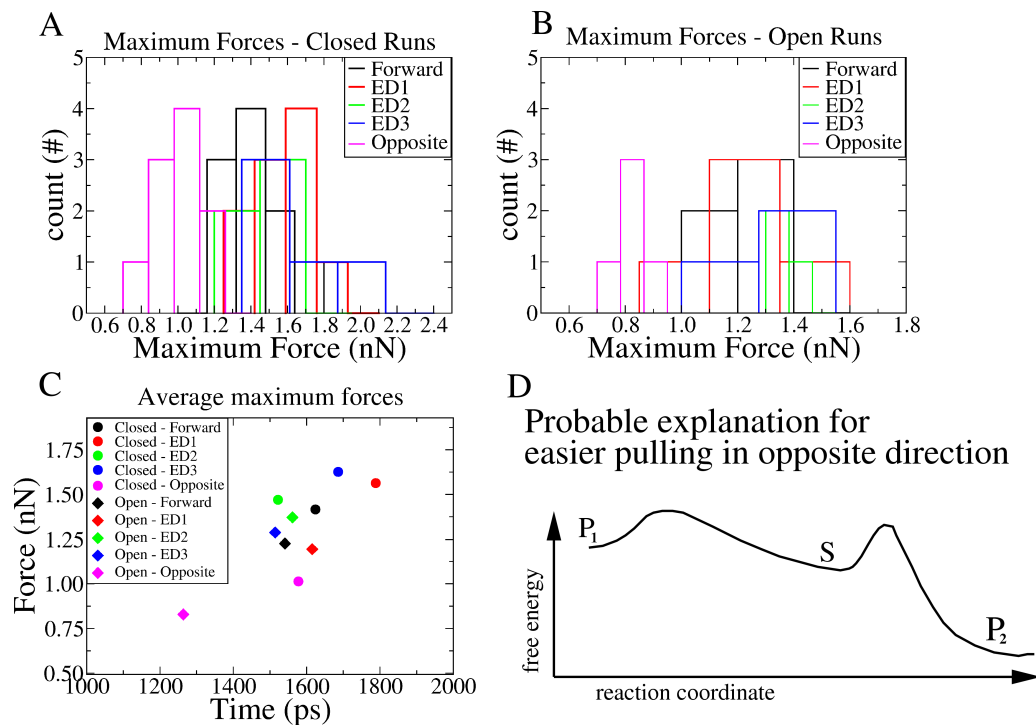
Figure 6.10: A and B: Histograms showing the maximum force rupture in all the pulling simulations discussed in Sec. 6.5.2. For the simulations starting with 'open' configuration the overall lower rupture forces show that opening of the cleft is favourable (or may be necessary) for the translocation. C: Average of the rupture forces in each case versus time corresponding to rupture event. All the simulations show faster translocation of the sugar for the 'open' set. D: Plausible explanation for the overall lower force profile for the pulling in opposite direction. The proposed energy barrier may have higher slope for the processive direction than for the non-processive direction.

Keeping the third eigenvector 'fixed' (which means the cleft access is locked) resulted in an apparent slight increase in the rupture force (blue curves). This might be correlated with the fact that the exit site has to open for the sugar translocation matching the observed projection of the trajectories onto this eigenvector. The increase in the projection relates to the opening of the cleft/access. This for the first time also strongly suggests a putative role of the third eigen-mode of the protein domain motions on the sugar translocation (in addition to the proposed role in the product-release mechanism).

Probably, the most interesting result is the lowest rupture forces required to pull the sugar in the opposite (non-processive) direction. This was very consistent in all the 15 simulations irrespective of whether the protein is 'closed' or 'open'. This low rupture force might indicate a favourable path for the sugar to move in the opposite direction and a higher energy barrier for the sugar to slide in the processive direction. The lower forces may indicate the presence of a significant energy barrier and in hence a more difficult forward pulling than pulling in the opposite direction. However, it is important

to note the difference between (pulling) forces and (free) energies. For example, in the scenario shown in Fig. 6.10 (D), if we start at S, in an FPMD simulation the maximum force towards P_2 might be higher than in the direction of P_1, whereas P_1 is higher in the free energy. The difference is due to the different slopes of the free energy profile.

The simulations described in this section gave some insight into the possible directionality of the processivity mechanism. The overall lower force profiles in the pulling simulations in the opposite direction can mean that some 'pushing' mechanism is required for the HA substrate to slide into the next round of catalysis. This may be important only for the short length HA substrates since for the longer HA-substrates the binding of the HA with the entry region (Fig. 5.11) along with the domain motions of the protein may play a role. However, it is difficult to address questions on the underlying free energy profile from such simulations. Therefore, we subsequently performed a series of simulations aimed at deriving an approximate free energy profile for the substrate sliding phase, as discussed in the next sections.

6.6   ED simulations: Results and discussions

As discussed earlier the pulling simulations provided valuable insight in the induced transition dynamics and it was seen from the observations in Sec. 6.5 that these simulations both confirmed the role of the protein dynamics on the mobility of the sugar and added a new finding concerning the directionality of the processive mechanism. However, the main disadvantage of the method is the fixed pulling direction. Essential Dynamics (ED) simulations may partly alleviate this bias by taking the reaction coordinate from the dynamics of the system itself, as deduced from the PCA on the free simulation.

To perform Essential dynamics (ED) simulations we took two structures: *Spn*Hyal with the hexasaccharide substrate shifted by one disaccharide unit (a snapshot from the equilibrated simulation HLT3water3, or sim. no. 10 of Chapter 5) and the x-ray structure 1loh. A PCA performed on these two structures yielded the reaction coordinate for the motion of the sugar by one disaccharide translocation (since that is the biggest difference in these two structures). Then with the EDSAM simulation technique the first structure was forced to attain the 1loh-like structure by slowly driving the system in the positive direction to sample this reaction coordinate. We carried out both the variants of the EDSAM technique as described in Sec. 6.4, *i.e.* forcing the system with fixed stepsize (linfix) or variable stepsize (linacc) along the chosen direction. Since the first four rings of the HA inside the cleft are from the crystal structure and hence — because of its helical nature — it was already rotated to some extent. This rotation was also reflected in the first eigenvector of the PCA. Thus, twisting of the HA was also induced as a part of the reaction coordinate, which may indicate that this reaction coordinate was indeed a more realistic coordinate than the one defined in the FPMD simulations.

As a first step we carried out two ED simulations with the 'linfix' method to force the sugar along the eigenvector within a fixed timeframe of 2 ns in the forward and the opposite direction. These simulations were performed on the structures where the
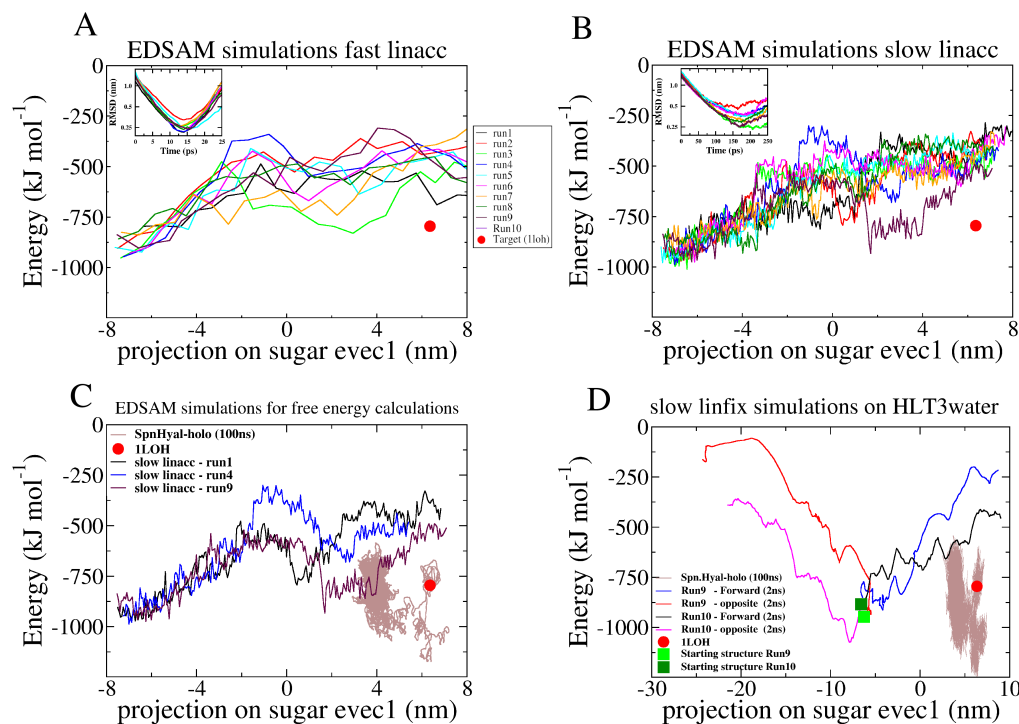
Figure 6.11: A and B: Essential Dynamics simulations performed for ten starting structures (the same as used for the FPMD simulations described in Sec. 6.5 with the 'closed' protein). Simulations without the stepsize control (A) and with stepsize control (B) are shown with the RMSD of the sugar in each simulation. The short range energy interactions of the sugar with the cleft as calculated in Sec. 5.3.6 provided a hint towards the proposed energy barrier (for example, green trace in A, maroon trace in B). C: Three representative cases from B were chosen for the calculations of the underlying free energy profile. D: Similar comparison for the EDSAM slow simulations (2 ns) with a fixed stepsize to drive the HA towards a 1loh-like position. Here, the system takes equal steps to achieve the target position and so is more 'biased' than the other simulations (as in A and B). The "forward" and "opposite" direction pulling show a comparable behaviour.

sugar was shifted by one disaccharide unit (one each from HLT3water1 (sim. no. 9) and HLT3water2 (sim. no. 10) of Chapter 5). The structures were taken from the equilibrated part of the respective simulation but when the sugar has yet not shown any processive motion towards 1loh-like position. The resulting sugar motion along with its short range interactions with the cleft are shown in Fig. 6.11 (D).

As described at the end of Sec. 6.4, we are interested in obtaining an approximate free energy profile for the sliding phase. To start simulations we first employed the 'linacc' method, in which the chosen speed can only serve as an upper limit parameter, to pull the sugar in the first set of pulling simulations (*i.e.* all the ten starting structures for which the protein was cleft was 'closed'). When the 'linacc' method was applied without any speed limit (*i.e.* when all spontaneous steps in the selected direction were accepted and all those in the opposite direction were rejected), all these simulations

showed a transition within 20 ps. This was surprisingly fast given the fact that no spontaneous transitions were observed in free simulations, even at the 100 ns timescale. Since 20 ps is extremely fast we performed the simulations with this set again with a slower speed of the translocation by providing upper limit for the stepsize of accepted steps. The simulations were then carried out for about 250 ps each. The trends observed were qualitatively similar to those observed from the behaviour in the fast simulations. Fig. 6.11 (B) shows the result for this set of simulations. Insets in each figure show the RMSD of the sugar for these simulations with respect to the target position of the sugar in 1LOH.pdb showing that the target (catalytic) position is reached within 2.5 Å RMSD. The sugar in both the cases moved very fast through the cleft along the chosen direction due to the applied bias. In all these simulations the translocation was acompanied by according protein domain motions (along the first three PCA eigenvectors of the protein) *i.e.* during these simulations the protein cleft opened and twisted as observed before in the FPMD simulations (not shown). Based on these energy-PCA curves (A and B), RMSDs of the sugar relative to 1loh position (insets for A and B), and protein domain motions (not shown), three simulations were identified as candidates for obtaining a free energy profile for the translocation of the sugar using a more extensive approach (see next section). The slow ED simulations for these simulations (run1, run4, and run9) are shown in Fig. 6.11 (C) where, after the sugar translocation, the simulations approached the regions that were spanned by the 'free' *Spn*Hyal-*holo* simulation (brown trace in C). The projections onto the PCA eigenvectors of the protein motions were also compared to 'free' *holo*- and *apo*-simulations.

6.7 Free energies

The free energy of a system is usually expressed as the Helmholtz function, A (for an NVT ensemble), or the Gibbs function, G (for an NPT ensemble), and is often considered as one of the most important quantities in thermodynamics because it determines the driving force for spontaneous processes. It is an equilibrium property of the system. Because of the sampling problem discussed earlier, the free energy calculation from a 'standard' MD simulation often poses a major challenge. However, making use of the fact that the free energy is a state function (independent of the pathways), it is possible to compute the free energy differences between a state A and a state B, $\Delta G = G^{\mathrm{B}} - G^{\mathrm{A}}$, by driving the system from A to B using the enforced techniques discussed earlier.

In the present work, approximate free energies for the *Spn*Hyal-HA system were calculated from equilibrium trajectories and from non-equilibrium trajectories. To derive free energies umbrella sampling techniques were applied and potentials of mean force were calculated from constraint forces applied to the sugar coordinate. For the non-equilibrium trajectories (e.g. FPMD) in principle the Jarzynski relation could be applied to derive a free energy profile (see also Sec. 6.7.3). In the present work we chose to focus primarily on the PMF and the umbrella sampling approach. These techniques and the preliminary results are outlined in the following sections Sec. 6.7.1 to 6.7.3.

6.7.1   Free energies from equilibrium trajectories

When FPMD simulations or ED simulations are carried out using a constraint force (in the harmonic sense this corresponds to an infinitely stiff spring), a free energy profile for the substrate translocation in the processive cycle can be constructed by integration over the mean force (averaged at several points along the pathway). Such a free energy profile (also termed potential of mean constraint force (PMF)), provides the information in which phase of the process barriers have to be overcome. Mapping to the structure then may tell which enzyme-substrate interactions play critical roles in that respect.

ED simulations (Sec. 6.4) force the system to move along the chosen principal modes. One simple method used in this thesis to estimate the free energy profile, was to fix the essential coordinates at n equidistant points, spaced by dz ('linfix' with stepsize 0). To obtain the equilibrium forces acting on the system at those coordinates the remaining $3N - 1$ degrees of freedom, for an N atom system, were allowed to equilibrate for a period of 10 to 25 ns for each set. Excluding the initial equilibration period, one can then calculate the mean force at each of these n points along the principal mode. The PMF is given by the sum over the average forces of the n points,

$$\Delta G_{\mathrm{PMF}} = \sum_n \langle F \rangle dz, \tag{6.4}$$

where dz is the distance between sampling points along the principal mode z, and $\langle F \rangle$ is the force acting on the system along the remaining coordinates.

Convergence can be slow due to equilibration effects and high force fluctuations. It can be assessed by comparing PMFs obtained from windows of different lengths and parts of the trajectories. Free energy profiles and convergence for the translocation of the HA-substrate in the *Spn*Hyal are discussed in Sec. 6.7.2.

Umbrella Sampling

Another widely used method to assess the free energy profile from equilibrated trajectories is *umbrella sampling*. In this method many overlapping windows along the chosen reaction coordinate are simulated simultaneously. An artificial harmonic biasing potential, called *umbrella potential* is added to each window simulation at systematically spaced positions along the reaction coordinate. This umbrella potential forces the system to compute an ensemble average over a biased distribution within that window. After sufficiently long simulations (equilibration per window) are carried out for each window, the umbrella bias is removed to obtain the final estimate of PMF.

The process of unbiasing and recombining the different simulation windows is the major difficulty in umbrella sampling and there are different strategies for it. One simple estimate is obtained from the mean deviation from the umbrella center per window, and integrating these mean forces. In this thesis, we used a more sophisticated strategy called *Weighted Histogram Analysis method*[32] (WHAM), which makes use of all the

information in umbrella sampling and does not discard the overlapping regions. In particular, the WHAM technique computes the total unbiased distribution function as a weighted sum of the unbiased distribution functions. The weighting function can be expressed in terms of known biased distribution functions. We use this procedure to obtain the PMF along the reaction coordinate of sugar-translocation obtained by PCA. At each window of simulation the bias potential was obtained by an inverted gaussian potential in the conformational flooding technique.[263]

6.7.2   Free energy profile for HA translocation

We employed the simulations to obtain a putative free energy profile for the HA translocation as described in the above two sections. As described at the end of Sec. 6.6, three simulations from slow EDSAM simulations were chosen for these calculations. To obtain the free energies as described in Sec. 6.7.1, 21 equidistant positions of the HA along the reaction coordinate were chosen for each of the three runs. By keeping the sugar 'locked' in each of these positions, the system was relaxed for a long time. The simulation time for each window was 25 ns (for run4), 10 ns (for run1) and 25 ns (for run9) thus the total simulation time was well above 1.2 μs. However, as described above, the convergence is very slow, presumably because of slow equilibration of perpendicular coordinates like the protein domain motions.

The free energy profiles are computed for every 250 ps window of the last 1.5 ns of run9 and are seen to slowly converge (see Fig. 6.12 A). From this profile a barrier of $\sim 10 - 20$ kJ/mol can be seen. Only 20 datapoints for each of these profile were available. Therefore, to obtain better statistics, the datapoints from all the simulations (run1, run4, and run9) were collected thus adding to at least 60 datapoints in total. The features of the energy profile did not change and are shown as the black curve in Fig. 6.12 (B). In both the profiles (computed from single run9 in Fig. 6.12 (A) and for all the datapoints taken together Fig. 6.12 (B), black curve), an energy barrier in the processive direction may correspond to the high forces required for the sugar to move in that direction in the FPMD simulations. To further validate this profile a set of 10 more 'free' or unbiased simulations (each of 10 ns) was performed with sugar at different intermediate positions around the region where an intermediate valley is seen (within 0 to -2 along the x-axis). The prjections of these free simulations are shown as overlayed histograms in Fig. 6.12 (A) (brown curves on the top of the profile). These histograms represent the probability distribution of the simulations along the reaction coordinate and were observed to quickly cluster around the intermediate minimum (around 0 on the $x-$axis) thereby confirming its presence. These simulations are all depicted by a red curve in Fig. 6.12 (B), which represents a very rough free energy estimate from the probability densities of these simulations according to the relation $G \approx -kTln\rho$ with ρ being the positional probability density along the reaction coordinate.

To obtain better statistics and better resolution of the free energy profile we performed umbrella sampling on two sets of the simulations (run4 and run9). In these simulations, structures at 60 equidistant windows for run9 and 50 equidistant windows
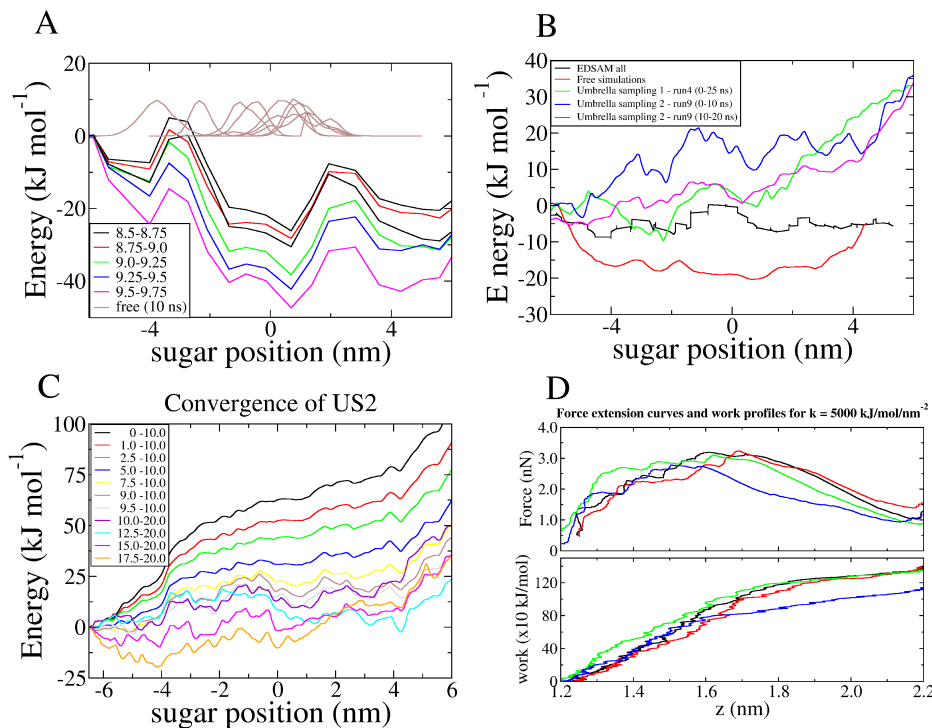
Figure 6.12: A: Free energy estimation carried out by employing Essential Dynamics simulations for the sugar translocation by restraining the HA at 21 different positions along the putative reaction coordinate. The energies are calculated for 250 ps windows of the last 1.5 ns of the total 10 ns trajectories for each position and the profile seems to converge slowly (black to magenta curves). The emergence of an intermediate minimum is validated by employing a series of free simulations (overlayed as brown histograms) each of 10 ns starting from different sugar positions along the reaction coordinate. The projections of these simulations to the sugar eigenvector show an increased probability distribution around this intermediate minimum and thereby confirm its presence. B: The umbrella sampling simulations performed for the two sets of runs (run4 and run9) from the earlier EDSAM simulations. The green curve shows the free energy profile obtained from the umbrella sampling for run4 simulating each of the 50 windows for 10 ns. The blue (first 10 ns) and magenta curves ($10 - 20$ ns) show similar estimates of the free energy from the simulations of 20 ns simulation of each of the 60 windows. The large deviation at the end is both due to poor overlap of the neighbouring umbrellas and the simulations being far from the equilibrium (since we drive the system fast with "linacc" simulations to this structure, long equilibration times are required). The red curve shows a rough estimate of the free energy profile extracted from the density probability from multiple free simulations using the relation $G \approx -kTln\rho$. The black curve shows the free energy estimation from the simulations of all the three sets of EDSAM simulations (run1, run4 and run9) with the sugar restrained at 21 different positions. C: Convergence for the umbrella sampling set US2. The profile converges fast in the initial periods (for the first 10 ns) and for the next 10 ns it is slowly converging. D: Force extension curves and work obtained for a very stiff spring as very first steps towards a free energy estimate via the Jarzynski Identity.

for run4 were obtained from the slow linacc simulations (Sec. 6.6). Umbrellas were created by biasing the simulations with a harmonic potential (as obtained by the inverse gaussian flooding[263]). The simulation was performed for 25 ns (for the first set US1 — with structures from run4) and 20 ns (for the second set US2 — with structures from run9) for each window thus totalling almost 2.2 μs. All the results taken together are shown in Fig. 6.12 (B). Here the green curve shows the free energy profile from the umbrella sampling on the first set (for 25 ns simulations). US1 showed convergence at least in the initial parts and the deviation at the right end was due to non-overlapping of the neighbouring umbrellas in that region posing a problem for the WHAM procedure. The free energy profile obtained in a similar manner for the second set is shown in blue (the first 10 ns) and magenta curves (for 10 to 20 ns). Umbrella sampling for this set is seen to be converging (from the blue curve for the first 10 ns to the magenta curve from 10 to 20 ns, see also Fig. 6.12 C) and is therefore still being continued for better statistics. Profiles from both sets at this stage, despite not being fully converged yet, confirm the features obtained from the earlier simulations, indicating a free energy barrier of about $\sim 10 - 20$ kJ/mol. The exact underlying cause of this energy barrier is still unknown and correlating back these profiles to the structures is planned as a future step of the analysis.

6.7.3 Free energies from non-equilibrium trajectories

It has recently been shown that free energy profiles can be derived from the irreversible work involved in non-equilibrium if the work is equal to or larger than the difference in free energy, $\Delta G \leq W$ simulations.[264–266]

Generally, for the transformation of a system from state A to state B under equilibrium conditions (for an infinitely slow transition) the work involved in this transition equals the difference in free energy, $\Delta G = W$. If the system is driven from state A to state B in finite time, *i.e.* under non-equilibrium conditions, part of the system energy is dissipated irreversibly as heat through friction. The work required for the transition is directly proportional to the force that is applied externally to induce the desired transition. In this case then, the work is larger than the difference in free energy, $\Delta G \leq W$.

For the general, non-equilibrium case the Jarzynski equality holds;

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G}. \tag{6.5}$$

This remarkable relation in principle allows to calculate free energy differences from non-equilibrium experiments and simulations.[266]

Thus, the free energies for any non-equilibrium simulations (in principle for an infinite number of A to B transitions) can be computed from the Jarzynski relation, saving computational efforts involved in 'free' or 'real-time' simulations. The free energy calculations based on the Jarzynski equation have been found to be comparable to those obtained by conventional equilibrium methods.[267] To calculate a free energy profile

using Eq. 6.5 from FPMD simulations for HA translocation in *Spn*Hyal we used an extremely stiff spring ($k = 5000$ kJ/mol/nm^2) and a very fast pulling speed (1 nm/100 ps). Fig. 6.12 (D) describes the work obtained from these simulations. However, it should be noted that to obtain the free energy profile comparable to potentials of mean force (PMF) obtained from the umbrella simulations described in the previous section, many sampling simulations are required (and also with much slower speed) and the calculation described here should only serve as proof of concept of how to employ this method. As can be seen from the umbrella sampling simulations (Fig. 6.12 (B)) each umbrella needs at least 20 ns (or more) for the protein to relax and hence to yield a reasonable estimate of the free energy. Since in the FPMD simulations as shown in Fig. 6.12 (D) the pulling speeds are too high ((1 nm/100 ps), the work obtained is also orders of magnitude higher, as compared to umbrella sampling indicating that an unrealistic pathway has been sampled. This also shows the limitation of the application of the Jarzynski relation with arbitrarily fast pulling simulations.

6.8 Conclusions

Mainly three advanced techniques to enforce the HA translocation were employed: high-temperature simulations, FPMD simulations, and ED sampling simulations. The simulations described in this chapter served two purposes. First, from the exhaustive 'free' or unbiased simulations, we have fair confidence that protein-ligand interplay occurs and in the active role of the protein domain motions in the processive mechanism. The enforced simulations described in this chapter fully confirmed these observations. These include mainly the domain motions of the protein and their effect on the translocation of the HA. This may be considered as a benchmark for these advanced techniques in which the system is no longer in equilibrium or 'free'. Second, with this benchmark, we performed extensive simulations to probe the putative free energy landscape involved in the processive mechanism. FPMD simulations provided the first hints on the shape of the landscape, and also provided insight into the involved roles of the domain motions of the protein on sugar translocation. From the sampling with this technique so far, convincing evidence was found that opening of the cleft favours (or may be even necessary) for the translocation of the sugar. Furthermore, some evidence was found that the other two domain motions (twisting and opening of the cleft exit) may play a role in sugar translocation but more sampling is required for better understanding of their role in the processive mechanism.

Finally, exhaustive simulations were performed to obtain an estimate for the underlying free energy profile for the sliding phase of the processive mechanism. These concur the previous observations, indicating a moderate free energy barrier in the processive direction that is low enough to be easily passed due to thermal energy, and can thus be expected to occur spontaneously. Based on the fact that no spontaneous translocation has been observed so far on the sub-microsecond timescale one may speculate that coupling to slow collective domain motions poses the rate limiting step for the translocation

part of the processive mechanism. An energy well is identified intermediate along the reaction pathway. This finding is also supported by a new set of 'free' simulations. The work involved in this chapter is still preliminary and extensive sampling in each method is required to further validate the results obtained.

Chapter 7

Epilogue

Science evolves not only by answering old questions but also by posing the new ones.
- Bert de Groot

In the introduction of this thesis, it was said that the purpose of the thesis is to address two different aspects of the processivity problem. First, the biophysical relevance of the processivity mechanism by studying the prototypic *Spn*Hyal-HA system, and to address the processivity mechanism in general broad perspective. Second, the fundamental knowledge of the molecular mechanism of the *Spn*Hyal system in particular details, which itself has physiological and medical importance. At the end, it is now the appropriate place to discuss the contributions of this thesis to the above mentioned fields with a short review of all the results.

The processivity mechanism is an inherently dynamic process and although the structural studies so far have given some hints towards the mechanism of processivity in many systems (like *Spn*Hyal, AP endonuclease), the main dynamical phase of how the substrates in these systems are actually translocated during multiple rounds of catalysis is difficult to study by experimental means. It was not clear until now, how this dynamical process of sliding in such systems can be explained only by the structural framework of the enzymes. We have tried to answer this longstanding question by elucidating the molecular mechanism at atomic detail by molecular dynamics simulations, which as the name suggests, is a technique developed to address dynamical problems. Indeed, with the particular example of the *Spn*Hyal-HA system, we are now able to suggest that only the knowledge of the structural framework of an enzyme is not sufficient to understand such a complex function. In fact, the processivity occurs as a very balanced conformational and energetic interplay between the enzyme and the substrate. The work addressed in this thesis has opened up many possibilities to study a variety of such processive enzymes in a biochemical or bio-engineering way and to add to general knowledge of this class of enzymes.

The sliding process generally requires a balance of the two opposite forces acting at the same time. A strong binding of the substrate to the enzyme is required for the substrate to remain attached to the enzyme, at the same time the binding must be weak enough so that the substrate can easily translocate or slide to accomplish the processivity. By employing various molecular dynamics simulations on the *Spn*Hyal-HA system (Chapter 5) we actually observed this change of the interactions at various stages of the processivity and a preliminary putative cause of the processivity could also be sketched. The processivity mechanism is in general much more complicated than just docking of the substrate to the enzyme, and the interplay between the conformational dynamics of an enzyme and the flexibility of the substrate may allow special interactions depending upon the system. Because of these interactions, it may be speculated that the processivity phase or the sliding phase may represent the main bottleneck to the time-limiting factor of the overall kinetics of the process. As an experimental means to validate this speculation, some AFM studies pulling the substrates in such a processive enzyme would help greatly to understand the scenario and mechanism underlying the sliding of the substrate. Simulation studies that address this issue at atomic detail are on the other hand very challenging, because the dynamic coupling between the enzyme and the substrate motions requires a large computational power for sufficient sampling of such events to have statistical confidence to speculate further on the underlying mechanism. With the ever increasing advent in computer technology it may yet be possible to study such systems more detail in coming years.

For the *Spn*Hyal-HA system, there are many promising findings that have indeed enriched the knowledge about the processivity mechanism in this system. The processivity mechanism in this system is mainly governed by the interplay of the domain motions of the protein that facilitates the sliding of the sugar. The twisting of the α-domain with respect to the β-domain has been speculated to be particularly important for the sliding mechanism. The sliding phase was found to occur at much longer timescales than expected and spontaneous processivity is still not accessible by the free simulations with the current state of the computer power. By employing various simulations, both 'free' and biased, we have been able to speculate about the principle time-limiting factor in this process. The investigation of the roles of the residues and charged regions of the cleft in various stages of the processivity mechanism - be it the sliding phase or the product release phase - have validated both the theoretical and experimental findings and also provided a wealth of knowledge to guide further studies along these lines.

Chapter 3 dealt with the comparison of different force fields and validated the approach of choosing the OPLS-AA force field (with some modifications in molecule specific dihedral terms). A comparison of the behaviour of the dynamics of the free sugar in aqueous solution and that of the sugar inside the cleft of *Spn*Hyal provided hints of different behaviour of the glycosidic dihedrals under both conditions. This work provided support to the previous finding (by x-ray crystallography, NMR, as well as other biophysical and computational methods like very short time MD) that hyaluronan is highly dynamic in aqueous solution. To validate the new force fields with these works and/or to gain any new insights into the long-term dynamics of the HA, we also have

started the studies of hyaluronan oligomers themselves in solution with a particular focus on aggregation. The simulations of different lengths of hyaluronan (2 to 10 rings) in aqueous solution were carried out to investigate the long term dynamics at microsecond timescales. The work may provide insights into the dynamical properties of this molecule at longer time scales.

Chapters 4 and 5 tell a story of how protein domain motions play an active role in the processive mechanism. In both these chapters extensive 'free' or unbiased simulations totalling to more than $\sim 1\mu$s have been presented. These simulations contain HA of different lengths inside the cleft at different phases of the mechanism (either in the processive phase or in the catalytic phase). From both the chapters important properties of the *Spn*Hyal that emerge are: the protein exhibits flexibility around the cleft region that may play a functional role in the processivity mechanism. Basically three major motional modes were identified from the PCA on these simulations: opening/closing of the protein cleft, twisting of the $\alpha-$domain (which contributes to almost all of the cleft) with respect to the $\beta-$domain, and the opening/closing of the entry/exit side of the cleft. All these motions were quantified and investigated for the relationship with each other and with respect to the sugar flexibility/motion inside the cleft. The opening/closing mode, which is also the largest amplitude mode, is found to be coupled to the twisting mode, the second largest amplitude mode, and these two seem to play a key role in the translocation of the sugar.

The positively charged cleft was proposed to cause strong binding for the negatively charged HA inside the cleft, but how the sugar can overcome these attractive forces while sliding through to the next catalytic round was unclear. In Chapter 5 we investigated the interactions of the sugar with the positive patch and an aromatic patch near the catalytic site as a whole and also on an individual residue level (for the residues that were known to be important from mutation experiments and some other residues in close vicinity of the sugar). From these investigations, it was seen that the positioning of the sugar inside the cleft was balanced by the interactions of the residues in the cleft positioned above and below the sugar. Interactions of HA with the aromatic patch differs depending upon whether the sugar is in the catalytic (largely attractive) or processive (sometimes repulsive interactions) phase and this may give a hint towards the role of this patch in the interplay of strong binding increasing the affinity within the cleft and weak binding essential for the sliding. More detailed investigations in this line (also complemented by the mutation experiments) would throw light into how the enzyme accomplishes these two seemingly opposite forces for the processive mechanism.

Although massive simulations of ~ 100 ns timescales were performed, the sugar in the simulations described in these Chapters (4 & 5) did not show spontaneous translocation by one complete disaccharide unit (the enzyme has been proposed to degrade the sugar processively mainly in disaccharide units[215]). Therefore, in subsequent studies we chose to apply an additional biasing potential during these simulations to be able to study enforced translocation in the expected direction. Chapter 6 which describes the enforced simulations focused on questions like what keeps the sugar not to translocate spontaneously on the sub-microsecond timescales, and whether there is any sign of an energy

barrier along the processive reaction coordinate. We performed first high-temperature simulations to increase the fluctuations and to see whether this minimal bias helps the sugar to translocate by one disaccharide unit. Although the sugar did not processively move, we saw strong correlation of the protein domain motions along the three domain motions that were studied in previous chapters. This reconfirmed the coupling between the dynamics of the protein and its function (the processivity mechanism).

Since the sugar did not move even when its temperatures was increased to as high as 750 K and since it is undesirable to increase the temperatures of the simulation systems to arbitrary high values, we then sought for other means to achieve translocation. In the first attempt, we performed a series of FPMD simulations to 'pull' the sugar through the cleft. With a reasonable force (spring constant $k = 750$ kJ/mol/nm^2) and pulling speed ($v = 1$ nm/ns) we achieved the desired translocation of the sugar for various situations. All these simulations concur to some extent with the 'free' simulations strongly supporting the role of domain motions of the protein coupling to the mobility of the sugar. Also the fact that these motions of the protein occur spontaneously as a response to the force induced on the sugar reflects the view that these motions are an inherent property of the protein. We convincingly observed that sugar slides with relative ease if the protein-cleft is open and/or kept open for the simulation time. Although it is intuitive that if the protein is open, the sugar is sliding with ease because of the obvious available larger space to fluctuate and move/slide, a very surprising result was obtained when the sugar was pulled in the opposite direction. In all the FPMD simulations, even a lower force was required for the sugar to slide in the reverse direction compared to the force required for the forward motion in any situation. This provided a hint for an energy barrier which may show different slopes in the different directions.

With this observation we went on to compute the free energy estimates for the processive mechanism. For this we employed the Essential Dynamics (ED) simulations and umbrella sampling techniques. The free energy calculations were found to require a large sampling and the computations are still continuing. However, the preliminary free energy estimates showed interesting features. Among these is a hint to an intermediate minimum which was confirmed by the 'free' simulations. Secondly, as expected, we indeed observe an energy barrier along the chosen reaction coordinate. Third, the energy barrier observed from the free energy profiles so far is not very high as compared to the thermal energy under physiological conditions so that they can be easily overcome by the thermal energy. The exact values of the energy barrier, its origin, the cross-relationship of the energy profile to the structural information and how the protein can overcome this barrier to accomplish the processivity are issues which need to be addressed in future. The low barrier heights may support the fact that the process can be expected to occur spontaneously. The fact that it does not within the 100 ns time scale may again indicate a strong coupling of sliding to the domain motions, which poses the rate limiting step for the translocation part of the processive mechanism of *Spn*Hyals.

7.1   what remains to be done...

In the end, like many PhD theses, I should also glorify the value of this work and tell stories of how it tried to solve many riddles of the processivity mechanism with the very specialised system, leading to answer the Question of Life, Universe, and Everything. But that would be a daringly false statement. Where does this thesis stand then? To put it in George B. Shaw's words "Science... never solves a problem without creating ten more." The studies described in this thesis have opened up many other questions related to the processivity mechanism of the *Spn*Hyal-HA system, and that, I would say, is the contribution of the thesis. There are a number of open questions at the moment. With new simulations along with experiments (especially site directed mutagenesis experiments) some of these questions may be answered in the future.

What can be a role simulations could play in further studies? Future steps would have to involve more sampling and more detailed analysis to address questions described below.

One possibility could involve simulations with neutral substrate or mutated enzyme. Simulations of artificially modified substrate analogs bound to Hyals may be carried out to investigate the role of specific interactions between the enzyme and the substrate. For example, the role of electrostatic interactions between the positively charged binding cleft of the protein and the negatively charged substrate can be investigated by monitoring the effect of computationally switching off the explicit charges on the substrate. Likewise, free energy perturbation (FEP) simulations may be carried out to model and analyse mutations within the hydrophobic patch of the enzyme. The latter have the advantage that not only the dynamic response to the perturbation can be monitored, but also the (binding) free energy cost (or gain) of the modification can be calculated.

From the present work the presence of an energy barrier has now become obvious. Apart from more analyses to probe the free energy estimates for the proposed mechanism, there is another interesting issue. It would be interesting to know how the enzyme accomplishes the directionality in the processive mechanism. How does it overcome the energy barrier in the pathways? In structural studies of the *Spn*Hyal the presence of a negatively charged region near the catalytic site of the cleft was observed. This led to the proposal of the repulsive mechanism for the negatively charged unsaturated disaccharide product. Leaving of the disaccharide product, possibly assisted by energy set free during catalysis, might therefore be the underlying mechanism that provides directionality to the processive mechanism. Some studies of this issue have already started and it will be interesting to study this aspect in detail in connection with the energetics of the free energy profile obtained. Both substrate and product may be placed inside the cleft and the release of the product (if it occurs on short time scales) may be investigated in detail for different lengths of the sugar.

From Chapter 4 and 5 the protein is seen to be mostly closed or 'locked' when the substrate is bound in the catalytic position, and to open up after the substrate has been removed. This can be revealed for many different lengths of the HA inside the cleft to

study the effect of substrate length on the processivity. Similar motions are expected in other bacterial Hyals. The overall dynamics would, however, have to be checked with multiple proteins of the Hyal family (including the mammalian hyaluronidases) in order to assess whether the findings observed here are transferable to other systems. These studies on one hand may give information about the common processivity mechanism of this family and on the other hand validate the simulations presented here. Also of interest are questions such as why *Spn*Hyal selectively degrades only HA and not other sugars? Are the domain motions observed in this thesis specific only when *Spn*Hyal interacts with HA or is it the general dynamical mechanism of these lyases for any substrate (Ch/ChS for example).

PCA however useful, is still restricted to only linear collective coordinates. It would be interesting to apply approaches that are not restricted to this limitation. One of these approaches would be Full Correlation Analysis method (FCA),[268] which was developed recently. Also, recently, an approach to assess general non-linear collective coordinates was developed by Schröder *et al.*,[269] which can be applied to detect the non-linearly correlated Hyal-HA dynamics.

The dynamics is happening at a much longer timescale than was presumed, and hence some sampling enhancement was applied like FPMD and EDSAM. These have to be justified with more sampling and also for various lengths of the HA. Some of the more advanced techniques like Jarzynski Identity[264, 265] (JI) or Crooks Fluctuation Theorem[270, 271] (CFT) can also be applied to these simulations to get the overall energy profile along the coordinate. Recently, a method named TeeREX[272] was developed by Kubitzki *et. al.* It would be interesting to apply this method to only selectively enhance the sampling along the sugar eigenvector in the replica exchange formalism.

Finally, and most importantly, the results obtained in the thesis have to be complimented or verified by experiments. Even after that, one can only hope that both the results from the simulations (which could be alleged to be 'unreal') and from the experiments (but still *in vitro*) correspond to the 'reality'. One can only pretend to know how the system is behaving under natural conditions and hope to apply the knowledge to the benefit of the society. After all, in Francis Crick's words[1]

> "...The main difficulty with large molecules, as opposed to small molecules, is that because they have so many parts the computation becomes almost prohibitively long. Nature's own analogue computer- the system itself-works so fantastically fast. Also she knows the rules more precisely than we do. But we still hope, if not to beat her at her game, at least to understand her — to give a concrete example, to calculate how a particular molecule folds itself up..."

Chapter 8

Musings...

I have learnt silence from the talkative, toleration from the intolerant, and kindness from the unkind; yet strange, I am ungrateful to these teachers.
- Kahlil Gibran (1883 - 1931)

Its just a beginning! It is a milestone! ... I can say I am quite thrilled at the end of this endeavour of thesis writing. After almost 3 months of collecting data and putting everything in context, I am right now full of many emotions. I am exhausted and in much need of sleep, I am satisfied that I was able to finish this thesis finally — although on quite a tight schedule, I am quite excited and motivated to do more work on the lose ends and unfinished stories of the hyaluronidases or to look into the problems in new perspectives, I am tensed to find some other position somewhere in the world and I am sad that my very cheerful days here in Göttingen are nearing their end. Yes, I am totally buried underneath the heap of thoughts and emotions, and if this is only because of the writing of this thesis, I now can (at least partly) understand how many authors feel when they finish the mammoth task of book-writing (of say a big textbook like Biochemistry[273] or a series of story books like Harry Potter) and why they express thanks to so many people. When I look back now, I cannot stop thinking of all the factors that helped me to reach this milestone in my life, and what a better place would be than here, to express the words of gratitude to all those who helped me in direct or indirect way to get here?

My first thanks is of course reserved to my supervisor Dr. Bert de Groot. Bert has always been there, and not like a boss, but mostly like a caring colleague. I am short of words in expressing my gratitude to him in helping me in every problem be it academic or personal. His way of looking positively at every single thing is amazing. There were times when I saw no light in this project and he always cheered me up (at least until the next wave of darkness) by showing the positive sides of my achievements until then. His patiently improving my soft skills in science, listening to each of my problems and

his ability to suggest a remedy for it, his vast experience and great enthusiasm towards research will always be my inspiration and to follow him suit is a goal to attain in my life.

The second most important person whom I want to thank is Evi Heinemann who helped me from the first day on. She has been very kind and compassionate all the time and in spite of her very busy schedule she always found time to help me get through German bureaucracy. Big thanks I owe also to Ingo Hoffmann, Oliver Slawik, Martin Fechner and Ansgar Esztermann for their enormous help by keeping all the computers in good health whenever we needed them.

I wish to thank Prof. Tim Salditt for his willingness to be the first referee for this thesis, and also all other members in the PhD committee.

I very much thank the whole department, its former and current members, for all the discussions, help with GROMACS and all the fun. Special thanks to Frauke Gräter who taught me FPMD simulations and Lars Schäfer for setting up QM simulations for HA parameters. Thanks to all my quake mates - Ben, Gunnar, Jochen, Martin (Stumpe), Maik, Marcus, Matthias, and also others with whom I had lot of fun whenever I was tired of trying to move the sugar. Thanks all you guys for making days in the lab enjoyable with all the cakes and icecream parties. Special thanks to Guillem Portella for his friendship and also for teaching me Linux, scripting and other details in the 'matrix world' in general. Discussions with you were always cheerful and informative and I will never forget your friendship. Thanks to Ulli, Olli, Ulf, and Marcus (Hennig) for staying up late at night in the lab and keeping the lab awake at nightshifts which helped me a lot during these years. Thanks to Juergen, Maik, Ben, and Guillem for making it easy to bridge the big cultural gap between me and the rest of the European world.

This thesis would not have been in this form if Ira had not read it and refurbished it from a novice diary to scientific account. Thanks Ira for your great and valuable help and constructive criticism. I know it was really horrible job for you to order the chaos from my writings.

The project would not have been fruitful without great help and support from Dr. Mark Jedrzejas. I am grateful to him for his hospitality during my visit to his lab in USA during the initial period of the project. I also wish to thank him for sparking my interest in molecular biology and biochemistry.

Thanks to MPG and DFG for financially supporting my stay here in Germany and allowing me to visit many conferences during my PhD. Thanks to MPI canteen for not letting me die of hunger (I really enjoyed the food despite what people accuse). I wish to thank the GROMACS developers (especially Carsten who helped me every now and then) and the GROMACS mailing list for enhancing technical knowledge about GROMACS. Thanks also to all developers and distributors of all the excellent softwares I have used.

I wish to thank all my Indian friends in Götingen and Germany especially Sunil, Nandan, Sadanand, Manmohan, Krishna, Anmol and Ritika, Avinash, and Debasis for

never letting me become homesick. My old friends Abhinav and Prasad (and his wife Anjali): thanks for all the good times visiting many places in and around Germany and feeding me great food whenever I visited you. I wish to thank all my friends: your friendship, support and never-failing encouragement has meant a lot to me, even if the possibilities to meet some of you did not materialise. Memory is always a handicap at such a moment when you are in dire need of it. I must have left out many names of people who have helped me and supported me willingly during all these years. I thank you all. I will always remember that without your timely help at the precise moments it would not be possible for me to be at this stage today.

Finally, it would not have been possible without constant suport of my parents and sisters who always believed in me more than myself. Mom and Dad, it is because of your backing me and granting me complete freedom to choose and follow this line of career that I am here today. I know you have sacrificed much for the sake of my studies. Thank you again for your love and support...

Appendix A

Abbreviations and Programs

A.1 Abbreviations commonly used

2D	two-dimensional
3D	three-dimensional
AFM	Atomic Force Microscopy
BTH	bovine testicular hyaluronidase
BVH	bee venom hyaluronidase
Ch(S)	Chondroitin(sulphate(s))
COM	center of Mass
DNA	deoxyribonucleic acid
DS	dermatan sulphate(s)
dsDNA	double-stranded deoxyribonucleic acid
ED	essential dynamics
FPMD	Force Probe Molecular Dynamics
fs	femto second
GAG(s)	glycosaminoglycan(s)
HA	Hyaluronan/Hyaluronic Acid
(h/)Hyal(s)	Hyaluronidase(s)
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
ns	nano second
PAD	Proton Acceptance and Donation
PCA	Principal Component Analysis
PMF	potential of mean force
ps	pico second
RMSD(s)	root mean square deviation(s)
RMSF(S)	root mean square fluctuation(S)
RNA	ribonucleic acid
S. or *Strep.*	*Streptococcus*
Spn	*Streptococcus pneumoniae*
ssDNA	single-stranded deoxyribonucleic acid
μs	micro second

A.2   Programs and softwares used

Protein Data Bank

All experimentally resolved biomolecular structures are deposited in a database called Protein Data Bank (PDB). The proteins in the database are identified by a unique four-letter identifier called PDB-ID. There are currently over 40 000 entries in PDB and is growing. The database can be accessed freely under at http://www.pdb.org.

The experimental conformations for *Spn*Hyal and for HA under study in this thesis were taken from this database. There are multiple models in case of structures resolved using NMR (in case of HA solution structures that were compared with the simulations to validate the modified parameter set of the HA molecule in Chapter 3. In these cases, for the sake of uniformity, we have always calculated RMSDs with respect to the first model in the PDB file.

GROMACS

All the MD simulations were carried out using GROMACS[60] simulation package.

http://www.gromacs.org

Gaussian

GAUSSIAN[274] is a software program popularly used for computational chemistry simulations. This program was used to calculate the partial charges for the HA parameter-set decision (see Appendix B for the details). http://www.gaussian.com/

Pymol and VMD

Visualisation softwares Pymol[275] and VMD[276] were used for visual representation of protein structures and the generation of high-resolution graphics for the thesis.

http://pymol.sourceforge.net/ ; http://www.ks.uiuc.edu/Research/vmd/

Xmgrace

All the graphs in this thesis were generated using Xmgrace which runs on unix-like systems. http://plasma-gate.weizmann.ac.il/Grace/

LATEX

The thesis was typesetted using LATEX in the Kile–1.8 integrated environment for LATEX.

http://www.latex-project.org/ ; http://kile.sourceforge.net/

Appendix B

HA force field parameters

Here we describe the parameter set (OPLS-AA*) used for the HA in all the simulations presented. This was derived from an improved version of the original OPLS-all atom (OPLS-AA) force field for carbohydrates.[277] The improvement was achieved by applying additional scaling factors for the electrostatic interactions between 1,5- and (in turn requiring scaling of) 1,6-interactions (for C4 and C6 atoms in hexapyranoses). This scaling was shown to improve the conformational energetics in comparison to the OPLS-AA force field. By applying such scaling for the electrostatic interactions of the 1,5- and some 1,6-interactions and by adjusting the torsional parameters we obtained a new set of parameters for HA. Futrthermore, the partial charges as obtained by quantum mechanical calculations were incorporated. The resulting set of parameters is shown to agree well with the available experimental structural data (Chapter 3) and is given in Table B.1 and B.3
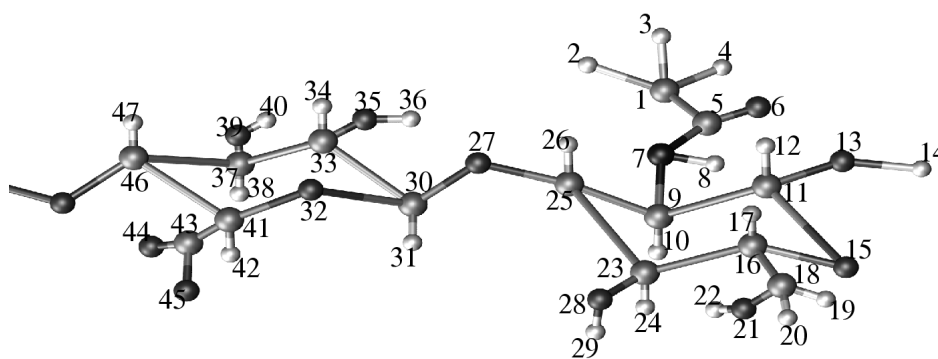
Figure B.1: schematic representation of HA disaccharide unit with the atom numbers labeled for the topology.

atom	atom type & opls name	charge	sigma[60]	epsilon[60]
	Non-bonded parameters			
1	CT (opls_135)	-0.319	3.50000e-01	2.76144e-01
2	HC (opls_140)	0.102	2.50000e-01	1.25520e-01
3	HC (opls_140)	0.102	2.50000e-01	1.25520e-01
4	HC (opls_140)	0.102	2.50000e-01	1.25520e-01
5	C (opls_235)	0.843	3.75000e-01	4.39320e-01
6	O (opls_236)	-0.654	2.96000e-01	8.78640e-01
7	N (opls_238)	-0.575	3.25000e-01	7.11280e-01
8	H (opls_241)	0.386	0.0	0.0
9	CT (opls_137)	-0.06	3.50000e-01	2.76144e-01
10	HC (opls_140)	0.06	2.50000e-01	1.25520e-01
11	CT (opls_183)	0.17	3.50000e-01	2.76144e-01
12	HC (opls_140)	0.06	2.50000e-01	1.25520e-01
13	OH (opls_169)	-0.437	3.07000e-01	7.11280e-01
14	HO (opls_170)	0.437	0.0	0.0
15	OS (opls_180)	-0.43	2.90000e-01	5.85760e-01
16	CT (opls_183)	0.17	3.50000e-01	2.76144e-01
17	HC (opls_185)	0.03	2.50000e-01	1.25520e-01
18	CT (opls_157)	0.309	3.50000e-01	2.76144e-01
19	HC (opls_140)	0.014	2.50000e-01	1.25520e-01
20	HC (opls_140)	0.009	2.50000e-01	1.25520e-01
21	OH (opls_154)	-0.760	3.12000e-01	7.11280e-01
22	HO (opls_170)	0.455	0.0	0.0
23	CT (opls_158)	0.246	3.50000e-01	2.76144e-01
24	HC (opls_140)	0.041	2.50000e-01	1.25520e-01
25	CT (opls_183)	0.269	3.50000e-01	2.76144e-01
26	HC (opls_185)	0.058	2.50000e-01	1.25520e-01
27	OS (opls_180)	-0.615	2.90000e-01	5.85760e-01
28	OH (opls_169)	-0.510	3.07000e-01	7.11280e-01
29	HO (opls_170)	0.510	0.0	0.0
30	CT (opls_183)	0.495	3.50000e-01	2.76144e-01
31	HC (opls_185)	0.014	2.50000e-01	1.25520e-01
32	CT (opls_158)	0.246	3.50000e-01	2.76144e-01
33	OS (opls_180)	-0.210	2.90000e-01	5.85760e-01
34	HC (opls_140)	0.078	2.50000e-01	1.25520e-01
35	OH (opls_169)	-0.756	3.07000e-01	7.11280e-01
36	HO (opls_170)	0.467	0.0	0.0
37	CT (opls_158)	0.210	3.50000e-01	2.76144e-01
38	HC (opls_140)	0.052	2.50000e-01	1.25520e-01
39	OH (opls_169)	-0.758	3.07000e-01	7.11280e-01
40	HO (opls_170)	0.497	0.0	0.0
41	CT (opls_183)	-0.163	3.50000e-01	2.76144e-01
contd on next page				

130

Non-bonded parameters continued from previous page				
atom	atom type & opls name	charge	sigma[60]	epsilon[60]
42	HC (opls_185)	0.096	2.50000e-01	1.25520e-01
43	C_3(opls_271)	0.993	3.75000e-01	4.39320e-01
44	O2 (opls_272)	-0.799	2.96000e-01	8.78640e-01
45	O2 (opls_272)	-0.815	2.96000e-01	8.78640e-01
46	CT (opls_183)	0.192	3.50000e-01	2.76144e-01
47	HC (opls_185)	0.057	2.50000e-01	1.25520e-01

Table B.1: Non-bonded parameters for the disaccharide unit of the HA.

Most of the bonded parameters are the same as the standard OPLS-AA parameters except a few molecule specific tosional parameters that were scaled. Regarding the scaling of these interactions, the torsional parametrs were changed according to the hexapyranoses terms as described by Kony et.al[76] (also given in Table B.2):

$$E_{torsion} = \sum_i \frac{V_1^i}{2}[1 + \cos(\theta + f_1)] + \sum_i \frac{V_2^i}{2}[1 - \cos(2\theta + f_2)] + \sum_i \frac{V_3^i}{2}[1 + \cos(3\theta + f_3)]$$

	OPLS–AA			OPLS–AA–SEI		
	V_1 [b]	V_2 [b]	V_3 [b]	V_1 [b]	V_2 [b]	V_3 [b]
Identical types						
C—C—C—C	1.740	−0.157	0.279	1.740	−0.157	0.279
C—C—C—H	0.000	0.000	0.366	0.000	0.000	0.366
H—C—O—C	0.000	0.000	0.760	0.000	0.000	0.760
C—O—C—C	0.650	−0.250	0.670	0.650	−0.250	0.670
H—C—C—O	0.000	0.000	0.468	0.000	0.000	0.468
H—C—C—H	0.000	0.000	0.318	0.000	0.000	0.318
H—C—O—H	0.000	0.000	0.450	0.000	0.000	0.450
Modified types						
O—C—O—C	−0.375	−1.358	0.004	−3.687	0.005	0.537
OS—C—C—OH	4.319	0.000	0.000	1.593	0.000	0.000
C—C—O—H	2.674	−2.883	1.026	0.205	−1.459	0.010
OH—C—C—OH	9.066	0.000	0.000	8.519	−6.829	4.793
O—C—O—H	−1.257	−1.806	0.003	−1.673	−2.200	0.001
C—C—C—OH	−1.336	0.000[c]	0.000[c]	−2.587	0.000	0.000
C—C—C—OS	−1.336	0.000[c]	0.000[c]	−0.089	−0.729	0.870
C—C—C—OS_ace	−1.336	0.000[c]	0.000[c]	1.501	0.000	0.000

[a]θ is the dihedral angle, and f_1, f_2, f_3 the phase shifts. All phase shifts are zero.
[b]Constants are given in kcal/mol.
[c]C—C—C—O is unique in OPLS–AA.

Table B.2: Sets of Tosional Angle parameters for Hexapyranoses for the OPLS-AA and OPLS-SEI force fields corresponding to the Potential Energy Function[a]. According to Kony et.al.

Accordingly, following terms were modified (the atom numbers are from Fig. B.1):

atoms in the torsion	the new OPLS–AA–SEI type
9 11 13 14	C—C—O—H
11 9 25 27	C—C—C—OS
15 16 23 25	C—C—C—OS
27 30 33 37	C—C—C—OS
32 30 33 37	C—C—C—OS
32 41 46 37	C—C—C—OS
43 41 46 48	C—C—C—OS
33 37 46 48	C—C—C—OS
23 16 18 21	C—C—C—OH
30 33 37 39	C—C—C—OH
39 37 46 41	C—C—C—OH
35 33 37 46	C—C—C—OH
30 33 35 36	C—C—O—H
37 33 35 36	C—C—O—H
33 37 39 40	C—C—O—H
46 37 39 40	C—C—O—H
46 41 43 44	C—C—C—OH
46 41 43 45	C—C—C—OH

Table B.3: List of torsional angles that were modified according to new OPLS-SEI force fields.

Appendix C

Principle of AFM experiments

Atomic force microscope (AFM) is probably one of the few instruments that currently provides atomic resolution and at the same time can be operated in solution. Because of these properties, AFM, originally developed as an imaging technique, is now also routinely used to monitor the dynamical changes in biomolecules as they 'work', i.e. during their functional cycle. During the years, major advances in optimisation in sample preparation,[278] image acquisition[279] and continuous instrumentation development[280] have made it possible now to manipulate single protein molecules and to measure the involved mechanical force.[281]
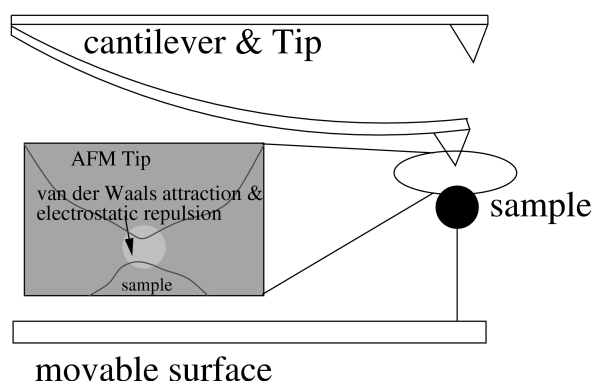
Figure C.1: A schematic representation of a typical AFM experiment. The sample molecule is attached to the movable surface. The cantilever is approached vertically until it makes contact with the other end of this sample. As the surface is moved slowly, the cantilever bends and this information can then be converted to to obtain the force-extension profiles yielding information about the mechanical properties of the system under study.

Fig. C.1 shows the schematic theme of a typical AFM single molecule experiment. The molecule under investigation is attached and kept fixed on one side to a surface *via* polymeric linkers. The cantilever or a flexible arm having a sharp tip at its end, is then approached vertically towards the sample until the other end of the molecule under study is attached to the tip. Now the surface is moved with a constant velocity thus

inducing a variation in the binding/unbinding force between the cantilever tip and the molecule. The force as a function of the cantilever position can be obtained from the subsequent bending of the cantilever. The resulting maximum forces (rupture forces) and the shape of the force profile can then give insights into the mechanical properties of the system, like unfolding/folding pathways, binding energies, or involved intermediates. Similar approaches that allow the mechanical manipulation of single molecules or cells are magnetic and laser tweezers[3] or the combinations of one or more of these.

Bibliography

[1] Francis Crick. *Of molecules and men.* University of Washington Press, 1967.

[2] C. R. Cantor and P. R. Schimmel. *Biophysical Chemistry - Part II: Techniques for the study of biological structure and function.* W.H.Freeman and Company, New York, 1997.

[3] G. Gompper, U. B. Kaupp, J. K. G. Dhont, D. Richter, and R. G. Winkler, editors. *Physics meets Biology - From Soft Matter to Cell Biology*, volume 19. Forschungszentrum Jülich GmbH, 2004.

[4] S. Weiss. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.*, 7:724–729, 2000.

[5] B. Schuler. Single-molecule fluorescence spectroscopy of protein folding. *Chem. Phys. Chem.*, 6:1206–1220, 2005.

[6] K. Besteman, S. Hage, N.H. Dekker, and S.G. Lemay. Role of tension and twist in single-molecule DNA condensation. *Phys. Rev. Lett.*, 98:058103, 2007.

[7] E. Z. Eisenmesser, D. A. Bosco, M. Akke, and D. Kern. Enzyme dynamics during catalysis. *Science*, 295:1520–1523, 2002.

[8] Y. Duan and P. Kollman. Pathways to a protein folding intermediate observed in a $\mu-$microsecond simulation in aqueous solution. *Science*, 282:740, 1998.

[9] J. Roth, F. Gähler, and H.R. Trebin. A molecular dynamics run with 5.180.116.000 particles. *Int.J. Mod. Phys. C*, 11:317–322, 2000.

[10] Essential dynamics of reversible peptide folding: Memory-free conformational dynamics governed by internal hydrogen bonds. B. l. de groot and x. daura and a. e. mark and h. grubmüller. *J. Mol. Biol.*, 390(1):299–313, 2001.

[11] R. Böckmann and H. Grubmüller. Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in f1-atp synthase. *Nat. Struct. Biol.*, 9:198–202, 2002.

[12] B. L. de Groot, G. Vriend, and H. J. C. Berendsen. Conformational changes in the chaperonin groel: New insights into the allosteric mechanism. *J. Mol. Biol.*, 286(4):1241–1249, 1999.

[13] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.

[14] M J. Jedrzejas. *Pneumococcal* virulance factors: Structure and function. *Microbiol. Mol. Rev.*, 65:187–202, 2001.

[15] P. H. V. Hippel, F. R. Fairfield, and M. K. Dolejsi. On the processivity of polymerases. *Ann. New York Acad. Sci.*, 726:118–131, 1994.

[16] K. Subramanian, W. Rutvisuttinunt, W. Scott, and R. S. Myers. The enzymatic basis of processivity in λ exonuclease. *Nuc. Acid Res.*, 31:1585–1596, 2003.

[17] C. Regni, A. M. Schramm, and L. J. Beamer. The reaction of phosphohexomutase from *pseudomonas aeruginosa. J. Biol. Chem.*, 281(22):15564–15571, June 2 2006.

[18] W. A. Breyer and B. W. Matthews. A structural basis for processivity. *Prot. Sci.*, 10:1699–1711, 2001.

[19] R. G. E. Coumans, J. A. A. W. Elemans, R. J. M. Nolte, and A. E. Rowan. Processive enzyme mimic: Kinetics and thermodynamics of the threading and sliding process. *Proc. Natl. Acad. Sci. USA*, 103(52):19647–19651, December 26 2006.

[20] K. Ponnuraj and M. J. Jedrzejas. Mechanism of hyaluronan binding and degradation: Structure of *streptococcus pneumoniae* haluronate lyase in complex with hyaluronic acid disaccharide at 1.7 å resolution. *J. Mol. Biol.*, 299:885–895, 2000.

[21] S. Li, K. B. Taylor, S. J. Kelly, and M. J. Jedrzejas. Vitamin C inhibits the enzymatic activity of *streptococcus pneumoniae* hyaluronate lyase. *J. Biol. Chem.*, 276(18):15125–15130, May 4 2001.

[22] M. J. Jedrzejas, L. V. Mello, B. L. de Groot, and S. Li. Mechanism of hyaluronan degradation by *streptococcus pneumoniae* hyaluronate lyase: Structures of complexes with the substrate. *J. Biol. Chem.*, 277:28287–28297, 2002.

[23] L. V. Mello, B. L. de Groot, S. Li, and M. J. Jedrzejas. Structure and flexibility of *streptococcus agalactiae* hyaluronate lyase complex with its substrate: Insight into the mechanism of processive degradation of hyaluronan. *J. Biol. Chem.*, 277:36678–36688, 2002.

[24] Z. Markovic-Housley, G. Miglierini, L. Soldatova, P. J. Rizkallah, U. Muller, and T. Schirmer. Crystal structure of hyaluronidase, a major allergen of bee venom. *Struct. Fold. Des.*, 8:1025–1035, 2000.

[25] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*, 102:6679–6685, 2005.

[26] J. Schlitter, M. Engels, and P. Kruger. Targeted molecular-dynamics - a new approach for searching pathways of conformational transitions. *J. Mol. Graph.*, 12:84–89, 1994.

[27] B. L. de Groot, A. Amadei, D. M. F. van Aalten, and H. J. C. Berendsen. Towards an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J. Biom. Str. Dyn.*, 13:741–751, 1996.

[28] A. Amadei, A. B. M. Linssen, B. L. de Groot, D. M. F. Aalten, and H. J. C. Berendsen. An efficient method for sampling the essential sub-space of proteins. *J. Biom. Str. Dyn.*, 13:615–626, 1996.

[29] H. Grubmüller. Predicting slow structural transitions in macromolcular systems - conformational flooding. *Phys. Rev. E*, 52:2893–2906, 1995.

[30] H. Grubmüller. Force probe molecular dynamics simulations. *Meth. Mol. Biol.*, 305:493–515, 2005.

[31] G. N. Patey and J. P. Valleau. Monte-carlo method for obtaining interionic potential of mean force in ionic solution. *J. Chem. Phys.*, 63:2334–2339, 1975.

[32] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules.1. the method. *J. Comp. Chem.*, 13:1011–1021, 1992.

[33] A. Laio and M. Parinello. Escaping free-energy minima. *Proc. natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[34] Sir Isaac Newton. *Philosophiae Naturalis Principia Mathematica (The Mathematical Principles of Natural Philosophy)*. London, July 5 1687.

[35] G. Gamow. *Thirty Years That Shook Physics: Story of Quantum Theory*. Dover Publications Inc., 1985.

[36] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford University Press, 1958.

[37] F. W. Byron and R. W. Fuller. *The Mathematics of Classical and Quantum Physics*. Dover Publications Inc., 1992.

[38] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, 1987.

[39] W. F. van Gunsteren, P.K. Weiner, and A. J. Wilkinson, editors. *Computer Simulation of Biomolecular Systems, theoretical and experimental applications.*, volume 3. Kluwer/ESCOM, 1997.

[40] W. F. van Gunsteren and H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29:992–1023, 1990.

[41] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 12:190–196, 2002.

[42] M.E. Tuckerman and G.J. Martyna. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B*, 104:159–178, 2000.

[43] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu. Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.*, 45:4064–4092, 2006.

[44] P. H. Hünenberger and W. F. van Gunsteren. *Computer Simulation of Biomolecular Systems, theoretical and experimental applications.*, volume 3, chapter Empirical classical interaction functions for molecular simulations, pages 3–82. Kluwer/ESCOM, 1997.

[45] A. R. Leach. *Molecular Modelling: Principles and Applications*. Addison Wesley Longman Limited, Essex, 1996.

[46] F. Jensen. *Computational Chemistry*. John Wiley and Sons Ltd, Sussex, England, 2001.

[47] G. Sutmann. *In Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, volume 10, chapter Classical Molecular Dynamics, pages 211–254. John von Neumann Institute for Computing (NIC), 2002.

[48] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys. Leipzig*, 84:457, 1927.

[49] P.W. Atkins and R. S. Friedman. *Molecular Quantum Mechanics*. Oxford University Press, 1996.

[50] M. Born and K. Huang. *Dynamical theory of crystal lattices*. Oxford University Press, 1954.

[51] F. C. Brooks. Convergence of intermolecular force series. *Phys. Rev.*, 86(1):92–97, April 1952.

[52] R. B. Gerber. On the order of accuracy of the born-oppenheimer approximation for molecular collision states. *Proc. Roy. Soc. London. Ser A - Math. and Phys. Sc.*, 309(1497):221–244, March 1969.

[53] A. Messiah. *Quantum Mechanics*, volume 2, chapter XVIII, pages 762–800. Dover Publications, 1999.

[54] L. Verlet. Computer "experiments" on classical fluids.i.thermodynamic properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[55] E. O. Kane. Theory of tunneling. *J. App. Phys.*, 32(1):83–91, January 1961.

[56] E. Gorham-Bergeron. Quantum mechanical theory of hydrogen diffusion. *Phys. Rev. Lett.*, 37(3):146–150, July 1976.

[57] A. Cohen and J. Klinman. Hydrogen tunneling in biology. *Chemistry & Biology*, 6(7):191–198, July 1999.

[58] L. Landau. Theory of the superfluidity of helium ii. *Phys. Rev.*, 60(4):356 – 358, August 1941.

[59] V. L. Ginzburg. Nobel lecture: On superconductivity and superfluidity (what I have and have not managed to do) as well as on the "physical minimum" at the beginning of the XXI century. *Rev. Mod. Phys.*, 76(3):981–998, July 2004.

[60] D. Van der Spoel, H. J. C. Berendsen, A. R. Van Buuren, E. Apol, P. J. Meulenhoff, A. L. T. M. Sijbers, and R. Van Drunen. *Gromacs User Manual.* Nijenborgh 4, 9747 AG Groningen, The Netherlands, 2006.

[61] J. A. McCammon, P.G. Wolynes, and M. Karplus. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry*, 18:927–942, 1979.

[62] M. Diraison, G. J. Martyna, and M. E. Tuckerman. Simulation studies of liquid ammonia by classical ab initio, classical, and path-integral molecular dynamics. *J. Chem. Phys.*, 111(3):1096–1103, 1999.

[63] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA*, 102:6665–6670, 2005.

[64] J. E. Lennard-Jones. *Proc. Roy. Soc. London Ser. A - Math. and Phys. Sc.*, 106:463, 1924.

[65] J. E. Lennard-Jones. On the forces between atoms and ions. *Proc. Roy. Soc. London Ser. A - Math. and Phys. Sc.*, 109:584–597, 1925.

[66] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.

[67] W. F. Gunsteren and H. J. C. Berendsen. *Gromos-87 manual.* BIOMOS b.v., Nijenborgh 4,9747 AG Groningen, The Netherlands., 1987.

[68] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.*, 7(2):230–252, 1986.

[69] W. L. Jorgensen and J. Tiraldo-Rives. The OPLS potential functions for proteins- energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[70] J. Tiraldo-Rives and W. L. Jorgensen. The OPLS force-field for organic and biomolecular systems. In *Abstr. Pap. Am. Chem. Soc., 204:43.* – COMP, Part 1, 1992.

[71] R. Lavery, K. Zakrzewska, and H. Sklenar. JUMNA (junction minimization of nucleic-acids). *Comput. Phys. Commun.*, 91(1-3):135–158, 1995.

[72] P. Derreumaux, M. Dauchez, and G. Vergoten. The structure and vibrational frequencies of a series of alkanes using SPASIBA force-field. *J. Mol. Struct.*, 295:203–221, 1993.

[73] K. Rasmussen. How to develop force fields: An account of the emergence of potential energy functions for saccharides. *J. Mol. Struct.*, 395-396:81–90, 1997.

[74] T. Herges and W. Wenzel. An all-atom forcefield for tertiary structure prediction of helical proteins. *Biophys. J.*, 87:3100, 2004.

[75] N. L. Allinger, Y. H. Yuh, and J. H. Lii. Molecular mechanics. the MM3 force fields for hydrocarbons. *J. Am. Chem. Soc.*, 111:8551–8566, 1989.

[76] D. Kony, W. Damm, and W. F. van Gunnsteren. An improved OPLS-AA force field for carbohydrates. *J. Comp. Chem.*, 23:1416–1429, 2002.

[77] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[78] W. L. Jorgensen and J. D. Madura. Temperature and size dependence for monte carlo simulations of tip4p water. *Mol. Phys.*, 56:1381–1392, 1985.

[79] M. Andresen, M. C. Wahl, A. C. Stiel, F. Gräter, L. V. Schäfer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmüller, S. W. Hell, , and S. Jakobs. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 102:13070–13074, 2005.

[80] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Advances in protein chemistry*, 66:27–85, 2003.

[81] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

[82] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7:306–317, 2001.

[83] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: Fast, flexible and free. *J. Comp. Chem.*, 26:1701–1718, 2005.

[84] R. W. Hockney. The potential calculation and some applications. *Meth Comput. Phys.*, 9:136–211, 1970.

[85] R. W. Hockney and J. W. Eastwood. *Computer Simulation using Particles*. Taylor & Francis, 1988.

[86] R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14:148–158, 1974.

[87] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. J. Tasumi. The protein data bank: A computer-based archival file for molecular structures. *J. Mol. Biol.*, 112:557, 1977.

[88] H. M. Berman, T. N. Bhat, P. E. Bourne, Z. Feng, G. Gilliland, H. Weissig, and J. West-brook. The protein data bank and the challenge of structural genomics. *Nat. Struct. Biol*, 7:957–959, 2000.

[89] MSI. *Insight II user guide*, October 1995.

[90] R. Koradi, M. Billeter, and K. Wüthrich. Molmol: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, 14(1):51–55, 1996.

[91] C. W. Gear. *Numerical initial value problems in ordinary differential equations*. Prentice Hall, Englewood Cliffs, NJ, 1971.

[92] W. C. Swope, H. C. Anderson, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76:637–649, 1982.

[93] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.*, 20:130–139, 1976.

[94] D. Janezic and B. Orel. Implicit runge-kutta methods for molecular dynamics integration. *J. Chem. Inf. Comput. Sci.*, 33:252–257, 1993.

[95] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2 edition, October 1992.

[96] E. Hairer, S. S. Norsett, and G. Wanner. *Solving Ordinary Dierential Equations I*. Springer-Verlag, New York/Berlin, 1987.

[97] Norbert Köckler. *Numerical Methods and Scientific Computing*. Clarendon Press, Oxford, 1994.

[98] M. P. Allen and D. J Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1987.

[99] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume 1 & 2. Addison Wesley, January 1989.

[100] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.

[101] M. F. Perutz. Electrostatic effects in proteins. *Science*, 201(4362):1187–1191, 1978.

[102] M. Orozco and F. J. Luque. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem Rev.*, 101(1):4187–4226, January 2001.

[103] D. Eisenberg and A. D. Mclachlan. Solvation energy in protein folding and binding. *Nature*, 319(6050):199–203, 1986.

[104] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[105] F. J. Luque, Y. Zhang, C. Aleman, M. Bachs, J. Gao, and M. Orozco. Solvent effects in chloroform solution: Parametrization of the MST/SCRF continuum model. *J. Phys. Chem.*, 100(10):4269–4276, 1996.

[106] B. Roux and T. Simonson. Implicit solvent models. *Biophys. Chem.*, 78:1–20, 1999.

[107] N. T. Southall, K. A. Dill, and A. D. J. Haymet. A view of the hydrophobic effect. *J. Phys. Chem. B*, 106(3):521–533, 2002.

[108] L. Wesson and D. Eisenberg. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Prot. Sci.*, 1:227–235, 1992.

[109] T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga. Accessible surface-areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. USA*, 84(10):3086–3090, 1987.

[110] C. A. Schiffer, J. W. Caldwell, R. M. Stroud, and P. A. Kollman. Inclusion of solvation free-energy with molecular mechanics energy – alanyl dipeptide as a test case. *Prot. Sci.*, 1(3):396–400, 1992.

[111] C. A. Schiffer, J. W. Caldwell, P. A. Kollman, and R. M. Stroud. Protein-structure prediction with a combined solvation free energy-molecular mechanics force-field. *Mol. Simul.*, 10(2-6):121, 1993.

[112] F. Fraternali and W. F. van Gunsteren. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. *J. Mol. Biol.*, 256(5):939–48, 1996.

[113] P. Ferrara, J. Apostolakis, and A. Caflisch. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins*, 46(1):24–33, 2002.

[114] A. Hiltpold, P. Ferrara, J. Gsponer, and A. Caflisch. Free energy surface of the helical peptide y(meara)(6). *J. Phys. Chem. B*, 104(43):10080–10086, 2000.

[115] K. Sharp. Incorporating solvent and ion screening into molecular-dynamics using the finite-difference poisson-boltzmann method. *J. Comp. Chem.*, 12(4):454–468, 1991.

[116] B. Z. Lu, W. Z. Chen, C. X. Wang, and X. J. Xu. Protein molecular dynamics with electrostatic force entirely determined by a single poisson-boltzmann calculation. *Proteins*, 48(3):497–504, 2002.

[117] M. K. Gilson and B. Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins: Struct. Funct. Genet.*, 4:7–18, 1988.

[118] Löeffer. Poisson-boltzmann calculations versus molecular dynamics simulations for calculating the electrostatic potential of a solvated peptide. *Theor. Chem. Acc.*, 101(1-3):163–169, 1999.

[119] R Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(Part I):255–284, 1966.

[120] F. Reif. *Fundamentals of Statistical and Thermal Physics.* McGraw-Hill, 1965.

[121] R. M. Levy and M. K. A. McCammon. Diffusive langevin dynamics of model alkanes. *Chem. Phys. Lett.*, 65:4–11, August 1979.

[122] R. Brustein, S. Marianer, and M. Schwartz. Langevin memory kernel and noise from lagrangian dynamics. *Physica A*, 175(1):47–58, June 1991.

[123] M. E. Tuckerman and B. J. Berne. Stochastic molecular dynamics in systems with multiple time scales and memory friction. *J. Chem. Phys.*, 95(6):4389–4396, 1991.

[124] T. Srokowski. Stochastic processes with finite correlation time: Modeling and applications to the generalized Langevin equation. *Phys. Rev. E.*, 64(3):031102, 2001.

[125] M. Berkowitz, J. D. Morgan, and J. A. McCammon. Generalized langevin dynamics simulations with arbitrary time-dependent memory kernels. *J. Chem. Phys.*, 78(6):3256–3261, March 15 1983.

[126] G. D. Harp and B. J. Berne. Time-correlation functions, memory functions, and molecular dynamics. *Phys. Rev. A*, 2(3):975, 1970.

[127] G. Otting, E. Liepinsh, and K. Wüthrich. Protein hydration in aqueous solution. *Science*, 254(5034):974–980, Nov 1991.

[128] Anna Rita Bizzarri and Salvatore Cannistraro. Molecular dynamics of water at the protein-solvent interface. *J. Phys. Chem. B*, 106(26):6617–6633, 2002.

[129] K. Bhattacharyya. Solvation dynamics and proton transfer in supramolecular assemblies. *Acc. Chem. Res.*, 36:95–101, 2003.

[130] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[131] M. Berkowitz and J. A. McCammon. Molecular dynamics with stochastic boundary conditions. *Chem. Phys. Lett.*, 90:215–217, 1982.

[132] A. J. Stace and J. N. Murrel. Molecular dynamics and chemical reactivity: A computer study of Iodine atom recombination under high pressure conditions. *Mol. Phys.*, 33:1–24, 1977.

[133] C. L. Brooks III and M. Karplus. Deformable stochastic bondary in molecular dynamics,. *J. Chem. Phys.*, 79:6312–6325, 1983.

[134] W. Im, S. Berneche, and B. Roux. Generalized solvent boundary potential for computer simulation. *J. Chem. Phys.*, 114:2924–2937, 2001.

[135] V. Gogonea. The solvent boundary potential: A new approach for computer simulation of large systems. *Internet Electron. J. Mol. Des.*, 2:179–194, 2003.

[136] W. Weber, P. H. Hünenberger, and J.A. McCammon. Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influenceof artificial periodicity on peptide conformation. *J. Phys. Chem. B*, 104:3668–3675, 2000.

[137] G. Gallavotti. *Statistical Mechanics - A short treatise.* Springer-Verlag, Berlin, 1999.

[138] R. K. Pathria. *Statistical Mechanics.* Butterworth-Heinemann, 1996.

[139] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.

[140] W. G. Hoover. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[141] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[142] W. F. van Gunsteren and M. Karplus. Effect of constraints on the dynamics of macro-molecules. *Macromolecules*, 15(6):1528–1544, 1982.

[143] S. Miyamoto and P. A. Kollman. SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water molecules. *J. Comp. Chem.*, 13:952–962, 1992.

[144] M. Saito. Molecular dynamics simulations of proteins in solution - artefacts by the cutoff approximation. *J. Chem. Phys.*, 101:4055–4061, 1994.

[145] M. Saito. Molecular dynamics simulations of proteins in solution with all degrees of freedom and long-range Coulomb interactions. *J. Phys. Chem.*, 99:17043–17048, 1995.

[146] P. P. Ewald. Die Berechung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, IV(64):253–287, 1920.

[147] T. Darden, D. York, and L. Pedersen. Particle mesh ewald - an N log N method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[148] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. The smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577, 1995.

[149] M. Levitt and S. Lifson. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.*, 46(2):269–279, 1969.

[150] A. T. Brunger and P. D. Adams. Molecular dynamics applied to x-ray structure refinement. *Acc. Chem. Res.*, 35(6):404–412, 2002.

[151] A. Kitao and N. Gō. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struc. Biol.*, 9(2):164–169, 1999.

[152] F. Tama, M. Valle, J. Frank, and C. L. Brooks III. Dynamic reorganization of the func-tionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA*, 100(16):9319–9323, August 5 2003.

[153] D.A. Case. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.*, 4:285–290, 1994.

[154] A. E. Garcia. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68(17):2696–2699, 1992.

[155] A. Amadei, A. B. M. Linsen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17(4):412–425, 1993.

[156] H. J. C. Berendsen and S. Hayward. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.*, 10(2):165–169, 2000.

[157] B. Brooks and M. Karplus. Harmonic dynamics of proteins - normal-modes and fluctua-tions in bovine pancreatic trypsin-inhibitor. *Proc. Natl. Acad. sci. USA*, 80(21):6571–6575, 1983.

[158] B. R. Brooks and M. Karplus. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. USA*, 82:4995–4999, 1985.

[159] N. Gō, T. Noguti, and T. Nishikawa. Dynamics of a small globular proteins in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*, 80:3696–3700, 1983.

[160] R. W. Harrison. Variational calculation of the normal modes of a large macromolecule: Methods and some initial results. *Biopolymers*, 23:2943–2949, 1984.

[161] W. Qian, J. Bandekar, and S. Krimm. Vibrational analysis of crystalline tri-l-alanine. *Biopolymers*, 31(2):193–210, 1991.

[162] P. G. Wolynes. Energy landscapes and solved protein-folding problems. *Phil. Trans. Roy. Soc. A*, 363(1827):453 – 467, February 15 2005.

[163] M. Karplus and J. N. Kushik. Method for estimating the configurational entropy of macro-molecules. *Macromolecules*, 14(2):325–332, 1981.

[164] A. Kitao, F. Hirata, and N. Gō. The effects of solvent on the conformation and the collective motions of protein - normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *chem. Phys.*, 158(2-3):447–472, 1991.

[165] R. M. Levy, M. Karplus, J. Kushick, and D. Perahia. Evaluation of the configurational entropy for proteins - application to molecular dynamics simulations of an alpha-helix. *Macromolecules*, 17(7):1370–1374, 1984.

[166] R. M. Levy, A. R. Srinivasan, W. K. Olson, and J. A. McCammon. Quasi-harmonic method for studying very low-frequency modes in proteins. *Biopolymers*, 23(6):1099–1112, 1984.

[167] M. M. Teeter and D. Case. Harmonic and quasiharmonic descriptions of crambin. *J. Phys. Chem.*, 94(21):8091–8097, 1990.

[168] A. L. Tournier and J. C. Smith. Principal components of the protein dynamical transition. *Phys. Rev. Lett.*, 91(20):208106, 2003.

[169] A. Amadei, B. L. de Groot, M. A. ceruso, M. Paci, A. Di Nola, and H. J. C. Berendsen. A kinetic model for the internal motions of proteins: Diffusion between multiple harmonic wells. *Proteins: Struct. Funct. Genet.*, 35:283–292, 1999.

[170] A. Kitao, S. Hayword, and N. Gō. Energy landscape of a native protein: Jumping-among-minima model. *Proteins: Struct. Funct. Genet.*, 33:496–517, 1998.

[171] M. A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100(7):2567–2572, 1996.

[172] A. Amadei, M. A. ceruso, and A. Di Nola. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins*, 36(4):419–424, 1999.

[173] T. S. Fisher, P. Joshi, and V. R. Prasad. Mutations thatconfer resistance to template-analog inhibitors of human immunodeficiency virus (HIV) type 1 reverse transcriptase lead to severe defects in HIV replication. *J. Virol.*, 76:291–299, 2002.

[174] E. Pascolo, C. Wenz, J. Linger, N. Hauel, H. Priepke, I. Hoffmann, P. Garin-Chesa, W. J. Rettig, K. Damm, and A. Schnapp. Mechanism of human telomerase inhibition by BIBR1532, a synthetic, non-nucleosidic drug candidate. *J. Biol. Chem.*, 277:15566–15572, 2002.

[175] S. B. Long, P.J. Casey, and L. S. Beese. Reaction path of protein farnesyltransferase at atomic resolution. *Nature*, 419:645–650, 2002.

[176] N. G. Nossal and M. F. Singer. Processive degradation of individual polyribonucleotide chains. I. *escherichia coli ribonuclease* 2. *J. Biol. Chem.*, 243:913–922, 1968.

[177] R. A. Bambara, D. Uyemura, and T. Choi. On processive mechanism of *escherichia coli* DNA-polymerase-I-quantitative assessment of processivity. *J. Biol. Chem.*, 253:413–423, 1978.

[178] R. A.. Bambara, P. J. Fay, and L. M. Mallaber. Methods of analysing processivity. *Methods Enzym.*, 262:270–280, 1995.

[179] D. G. Pritchard, J. O. Trent, X. Li, P., Zhang, and M. L. Egan. Characterization of the active site of group b streptococcal hyaluronan lyase. *Proteins*, 40:126–134, 2000.

[180] W. L. Hynes and S. L. Walton. Hyaluronidases of gram-positive bacteria. *FEMS Microbiol Lett*, 183:201–207, 2000.

[181] N. Kostyukova, M. Volkova, V. Ivanova, and A. Kvetnaya. A study of pathogenic factors of streptococcus pneumoniae strains causing meningitis. *FEMS Immunol. Med. Microbiol.*, 10:133–137, 1995.

[182] F. Duran-Reynals. Exaltation de l'activité du virus vaccinal par les extraits de certains organes. *CR Séances Soc. Biol Fil.*, 99:6–7, 1928.

[183] K. Meyer, R. Dubos, and E. M. Smith. The hydrolysis of the polysaccharide acids of vitreous humor, of umbilical cord and of streptococcus by the autolytic enzyme of pneumococcus. *J. Biol. Chem.*, 118(1):71–78, 1937.

[184] E. Chain and E. S. Duthie. A mucolytic enzyme in testis extracts. *Nature*, 144:977–78, 1939.

[185] G. Lepperdinger, B. Strobl, and G. Kreil. Hyal2, a human gene expressed in many cells, encodes a lysosomal hyaluronidase with a novel type of specificity. *J. Biol. Chem.*, 273(35):22466 – 22470, August 28 1998.

[186] T. Isoyama, D. Thwaites, M. G. Selzer, R. I. Carey, R. Barbucci, and V. B. Lokeshwar. Differential selectivity of hyaluronidase inhibitors toward acidic and basic hyaluronidases. *Glycobiology*, 16(1):11 – 21, January 1 2006.

[187] A. B. Csóka, , and R. Stern. Mammalian hyaluronidases. In V. C. Hascall and M. Yanagishita, editors, *Science of Hyaluronan Today*. Glycoforum, 2000.

[188] G. L. Frost, T. Csoka, and R. Stern. The hyaluronidases: a chemical, biological and clinical overview. *Trends Glycosci. Glycotechnol.*, 8:419–434, 1996.

[189] E. Kim, D. Baba, M. Kimura, M. Yamashita, S. i. Kashiwabara, and T. Baba. Identification of a hyaluronidase, hyal5, involved in penetration of mouse sperm through cumulus mass. *Proc. Natl. Acad. Sci. USA*, 102(50):18028 – 18033, December 13 2005.

[190] G. Kreil. Hyaluronidases–a group of neglected enzymes. *Prot. Sci.*, 4:1666–1669, 1995.

[191] RO Miura, S. Yamagata, Y. Miura, T. Harada, and T. Yamagata. Analysis of glycosaminoglycan-degrading enzymes by substrate gel electrophoresis (zymography). *Anal Biochem*, 225:333–340, 1995.

[192] A. B. Csoèkaa, G. I. Frost, T. Wong, and R. Stern. Purifiation and microsequencing of hyaluronidase isozymes from human urine. *FEBS Lett.*, 417(3):307–310, 1997.

[193] E.J. Menzel and C. Farr. Hyaluronidase and its substrate hyaluronan: biochemistry, biological activities and therapeutic uses. *Cancer Lett.*, 131(1):3–11, September 11 1998.

[194] Y. Lin, K. Mahan andW. F. Lathrop, D. G. Myles, and P. Primakoff. hyaluronidase activity of the sperm plasma membrane protein ph-20 enables sperm to penetrate the cumulus cell layer surrounding the egg. *J Cell Biol*, 125:1157–1163, 1994.

[195] M. A. Zahalka, E. E. Okon, U. Gosslar, B. B. Holzmann, and D. Naor. Lymph node (but not spleen) invasion by murine lymphoma is both cd44- and hyaluronate- dependent. *J Immunol*, 154:5345–5355, 1995.

[196] T. Spruss, G. G. Bernhardt, and H. Schonenberger W. Schiess. Hyaluronidase significantly enhances the efficacy of regional vinblastine chemotherapy of malignant melanoma. *J Cancer Res Clin Oncol*, 121:193–202, 1995.

[197] A. Dorfman, M. L., and R. Whitney. The hyaluronidase inhibitor of human blood. *J Biol. Chem.*, 174:621–629, 1948.

[198] U. R. Kuppusamy, H. E. Khoo, and N. P. Das. Structure-activity studies of flavonoids as inhibitors of hyaluronidase. *Biochem Pharmacol*, 40:397–401, 1990.

[199] K. Mio and R. Stern. Inhibitors of the hyaluronidases. *Matrix Biol.*, 21:31–37, 2002.

[200] K. Suzuki, Y. Terasaki Y, and M. Uyeda. Inhibition of hyaluronidases and chondroitinases by fatty acids. *J Enzyme Inhib Med Chem*, 17:183–186, 2002.

[201] M. S. Akhtar and V. Bhakuni. Streptococcus pneumoniae hyaluronate lyase contains two non-cooperative independent folding/unfolding structural domains: characterization of functional domain and inhibitors of enzyme. *J. Biol. Chem.*, 278:25509–25516, 2003.

[202] S. Braun. *New Inhibitors of bacterial hyaluronidase - Synthesis and structure-activity relationships*. PhD thesis, Universität Regensburg, 2005.

[203] K. Meyer. *Hyaluronidases: The Enzymes.* Academic Press, New York, 1971.

[204] K. Takagaki, T. Nakamura, J. Izumi, H. Saitoh, and M. Endo. Characterization of hyadrolysis and transglycosylation by testicular hyaluronidase using ion-spray mass spectroscopy. *Biochem.*, 33:6503–6507, 1994.

[205] S. Suzuki. Microbial hyaluronan lyases. 2000.

[206] A. B. Csoka, G. I. Frost, and R. Stern. The six hyaluronidase-like genes in the human and mouse genomes. *Matrix Biol.*, 20:499–508, 2001.

[207] R. Stern. Devising a pathway for hyaluronan catabolism: are we there yet? *Glycobiology*, 13:105R–115R, 2003.

[208] P. M. Coutinho and B. Henrissat. *Recent Advances in Carbohydrate Bioengineering*, chapter Carbohydrate-active enzymes: an integrated database approach, pages 3–12. The Royal Society of Chemistry, 1999.

[209] M.J. Jedrzejas and L. Chantalat. Structural studies of *streptococcus agalactiae* hyaluronate lyase. *Acta Crystallogr. D. Biol. Crystallogr.*, 56:460–463, 2000.

[210] M. J. Jedrzejas. Three-dimensional structures of hyaluronate lyases from *streptococcus* species and their mechanism of hyaluronan degradation. In V. C. Hascall and M. Yanagishita, editors, *Science of Hyaluronan Today.* Glycoforum, www.glycoforum.gr.jp/science/hyaluronan, 2002.

[211] S. Li, S. J. Kelly, E. Lamani M. Ferraroni, and M. J. Jedrzejas. Structural basis of hyaluronan degradation by *streptococcus pneumoniae* hyaluronate lyase. *EMBO J*, 19:1228–1240, 2000.

[212] S. Li and M. J. Jedrzejas. Hyaluronan binding and degradation by *streptococcus agalactiae* hyaluronate lyase. *J. Biol. Chem.*, 276:41407–41416, 2001.

[213] M. J. Jedrzejas, L. Chantalat, and R. B. Mewbourne. Crystallization and preliminary x-ray analysis of *streptococcus pneumoniae* hyaluronate lyase. *J Struct. Biol.*, 121:73–75, Jan 1 1998.

[214] M. J. Jedrzejas, R. B. Mewbourne, L. Chantalat, and D. T. McPherson. Expression and purification of *streptococcus pneumoniae* hyaluronate lyase from *escherichia coli*. *Protein Expr. Purif.*, 13(1):83–89, Jun 1998.

[215] S. J. Kelly, K. B. Taylor, S. Li, and M. J. Jedrzejas. Kinetic properties of *streptococcus pneumoniae* hyaluronate lyase. *Glycobiology*, 11(4):297–304., Apr 2001.

[216] M. Nukui, K. B. Taylor, D. T. McPherson, M. K. Shigenaga, and M. J. Jedrzejas. The function of hydrophobic residues in the catalytic cleft of *streptococcus pneumoniae* hyaluronate lyase. kinetic characterization of mutant enzyme forms. *J Biol Chem.*, 278(5):3079–3088, Jan 31 2003.

[217] D. J. Rigden and M. J. Jedrzejas. Genome-based identification of a carbohydrate binding module in *streptococcus pneumoniae* hyaluronate lyase. *Proteins*, 52(2):203–211, Aug 1 2003.

[218] D. J. Rigden and M. J. Jedrzejas. Structures of *streptococcus pneumoniae* hyaluronate lyase in complex with chondroitin and chondroitin sulfate disaccharides. insights into specificity and mechanism of action. *J Biol Chem.*, 278(50):50596–50606, Dec 12 2003.

[219] D. J. Rigden, J E. Littlejohn, H. V. Joshi, B. L. de Groot, and M. J. Jedrzejas. Alternate structural conformations of streptococcus pneumoniae hyaluronan lyase: insights into enzyme flexibility and underlying molecular mechanism of action. *J. Mol. Biol.*, 358(4):1165–78, May 12 2006.

[220] D. J. Rigden, A. Botzki, M. Nukui, R. B. Mewbourne, E. Lamani, S. Braun, E. von Angerer, G. Bernhardt, S. Dove, A. Buschauer, and M. J. Jedrzejas. Design of new benzoxazole-2-thione derived inhibitors of streptococcus pneumoniae hyaluronan lyase: Structure of a complex with a 2-phenylindole. *Glycobiology*, 16(8):757–765, 2006.

[221] Md. S. Akhtar, M. Y. Krishnan, and V. Bhakuni. Insights into the mechanism of action of hyaluronate lyase: Role of c-terminal domain and Ca^{2+} in the functional regulation of enzyme. *J. Biol. Chem.*, 281(38):28336–28344, 2006.

[222] G. J. Boulnois. Pneumococcal proteins and the pathogenesis of disease caused by *streptococcus pneumoniae*. *J. Gen. Microbiol.*, 138:249–259, 1992.

[223] B.L. de Groot, D.M.F. van Aalten, R.M. Scheek, A. Amadei, G. Vriend, and H.J.C. Berendsen. Prediction of protein conformational freedom from distance constraints. *Proteins*, 29:240–251, 1997.

[224] K. Meyer and J. W. Palmer. The polysaccharide of the vitreous humor. *J. Biol. Chem.*, 107:629–634, 1934.

[225] B. Weissman and K. Meyer. The structure of hyalobiuronic acid and of hyaluronic acid from umbilical cord. *J. Am. Chem. Soc.*, 76:1753–1757, 1954.

[226] E. A. Balasz, T. C. Laurent, and R. W. Jeanloz. Nomenclature of hyaluronic acid. *Biochem. J.*, 235:903, 1986.

[227] V. Hascall and T. C. Laurent. Hyaluronan: Structure and physical properties. 2000.

[228] J. E. Scott. Secondary structures in hyaluronan solutions: chemical and biological implications. *The Biology of Hyaluronan. Ciba Foundation Symposium*, 143:6–14, 1989.

[229] J. E. Scott, C. Cummings, A. Brass, and Y. Chen. Secondary and tertiary structures of hyaluronan in aqueous solution, investigated by rotary shadowing-electron microscopy and computer simulation. hyaluronan is a very efficient network-forming polymer. *Biochem. J.*, 274:699–705, 1991.

[230] A. Almond, P. L. DeAngelis, and C. D. Blundell. Hyaluronan: the local solution conformation determined by nmr and computer modelling is close to a contracted left-handed four-fold helix. *J. Mol. Biol.*, 358:1256–69, 2006.

[231] Glycoforum http://www.glycoforum.gr.jp/, 2005.

[232] J. E. Scott and F. Heatley. Hyaluronan forms specific stable tertiary structures in aqueous solution: a 13c nmr study. *Proc. Natl. Acad. Sci. USA*, 96:4850–4855, 1999.

[233] J. E. Scott and F. Heatley. Biological properties of hyaluronan in aqueous solution are controlled and sequenced by reversible tertiary structures, defined by nmr spetroscopy. *Biomacromolecules*, 3:547–553, 2002.

[234] T. C. Laurent and J. R. Fraser. Hyaluronan. *FASEB J*, 6:2397–2404, 1992.

[235] T. C. Laurent, U. B. Laurent, and J. R. Fraser. The structure and function of hyaluronan: An overview. *Immunol. Cell. Biol.*, 74:A1–7, 1996.

[236] T. B. Csoka, G. I. Frost, and R. Stern. Hyaluronidases in tissue invasion. *Invasion Metastasis*, 17:297–311, 1997.

[237] D. C West and H.Chen. Is hyaluronan degradation an angiogenic/metastatic switch? *New Frontiers in Medicinal Sciences: Redefining Hyaluronan, Elsevier Science*, pages 77–86, 2000.

[238] A. J. Day and G. D. Prestwich. Hyaluronan-binding proteins: tying up the giant. *J. Biol. Chem.*, 277:4585–4588, 2002.

[239] D. G. Jackson. The lymphatic endothelial hyaluronan receptor lyve-1. In *http://www.glycoforum.gr.jp/science/hyaluronan/HA28/HA28E.html*, 2004.

[240] K. Kimata and L. Zhuo. Shap, a protein covalently bound to hyaluronan. In *http://www.glycoforum.gr.jp/science/hyaluronan/HA22/HA22E.html*, 2001.

[241] G. Frost, T. Csoka, and R. Stern (1996). The hyaluronidases: a chemical, biological and clinical overview. *Trends Glycosci Glycotechnol.*, 8:419–434, 1996.

[242] M. R. Natowicz, M. P. Short, Y. Wang, G. R. Dickersin, M. C. Gebhardt, D. I. Rosenthal, K. B. Sims, and A. E. Rosenberg. Clinical and biochemical manifestations of hyaluronidase deficiency. *N Engl J Med*, 335:1029–33, 1996.

[243] E. A. Balazs and J. L. Denlinger. Clinical uses of hyaluronan. *Ciba Found Symp*, 143:265–75; discussion 275–80, 281–5, 1989.

[244] K. Fukuda, H. Dan, M. Takayama, F. Kumano, M. Saitoh, and S. Tanaka. Hyaluronic acid increases proteoglycan synthesis in bovine articular cartilage in the presence of interleukin-1. *J Pharmacol Exp Ther*, 277:1672–5, 1996.

[245] A. Asari, S. Miyauchi, S. Matsuzaka, T. Ito, E. Kominami, and Y. Uchiyama. Molecular weight-dependent effects of hyaluronate on the arthritic synovium. *Arch Histol Cytol*, 61:125–35, 1998.

[246] S. Gotoh, J. Onaya, M. Abe, K. Miyazaki, A. Hamai, K. Horie, and K. Tokuyasu. Effects of the molecular weight of hyaluronic acid and its action mechanisms on experimental joint pain in rats. *Ann Rheum Dis*, 52:817–22, 1993.

[247] H. Iwata. Pharmacologic and clinical aspects of intraarticular injection of hyaluronate. *Clin. Orthop.*, pages 285–91, 1993.

[248] R. D. Lins and P. H. Hünenberger. A new gromos force field for hexopyranose-based carbohydrates. *J. Comp. Chem.*, 26:1400–1412, 2005.

[249] R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM-93 parameter development. *J. Phys. Chem.*, 99:3832–3846, 1995.

[250] D. Kony, W. Damm, and W. F. Gunsteren. An improved opls-aa force field for carbohydrates. *J. Comp. Chem.*, 23:1416–1429, 2002.

[251] B. Heymann and H. Grubmüller. 'chair-boat' transitions and side groups affect the stiffness of polysaccharides. *Chem. Phys. Lett.*, 305:202–208, 1999.

[252] D. N. Wilson, F. Schlünzen, M J. Harms, T. Yoshida, T. Ohkubo, R. Albrecht, J. Buerger, Y. Kobayashi, and P. Fucini. Ribosome recycling and termination of protein synthesis. *EMBO J.*, 24(2):251–260, 2005.

[253] G. Vriend. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[254] J. E. Scott and A. M. Thomlinson. The structure of interfibrillar proteoglcan bridges (shape modules') in extracellular matrix of fibrous onnective tissues and their stability in various chemical environment. *J. Anat.*, 192:391–405, 1998.

[255] R. E. Turner, P. Y. Lin, and M. K. Cowman. Self-association of hyaluronate segments in aqueous nacl solution. *Arch. Biochem. Biophys.*, 265:484–495, 1998.

[256] K. Tai. Conformational sampling for the impatient. *Biophys. Chem.*, 107:213–220, 2004.

[257] A. DiNola, D. Roccatano, and H. J. C. Berendsen. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins: Struct. Funct. Genet.*, 19:174–182, 1994.

[258] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11(24):230, 2001.

[259] F. Gräter, J. H. Shen, H. L. Jiang, M. Gautel, and H. Grubmüller. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.*, 88:790–804, 2005.

[260] H. Grubmüller, B. Heynmann, and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271:997–999, 1996.

[261] F. Gräter. *Enforced functional protein motions studied by molecular dynamics simulations.* PhD thesis, Max Planck Institute for Biophysical Chemistry, Göttingen, 2005.

[262] B. Heymann and H. Grubmüller. Dynamic force spectroscopy of molecular adhesion bonds. *Phys. Rev. Lett.*, 84:6126–6129, 2000.

[263] O.F. Lange, L.V. Schäfer, and H. Grubmüller. Flooding in gromacs: Accelerated barrier crossings in molecular dynamics. *J. Comp. Chem.*, 27:1693–1702, 2006.

[264] C. Jarzynski. Nonequilibrium equality for free energy difference. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[265] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018–5035, 1997.

[266] G. Hummer and A. Szabo. Free energy reconstruction from non-equilibrium single-molecule pulling experiments. *Proc. natl. Acad. Sci. USA*, 98:3658–3661, 2001.

[267] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *j. Chem. Phys.*, 119:3559–3566, 2003.

[268] O. Lange. *Collective Langevin Dynamics of Conformational Motions in Proteins*. PhD thesis, Göttingen, 2005.

[269] G.F. Schröder. *Simulation of Fluorescence Spectroscopy Experiments*. PhD thesis, Universität Göttingen, 2004.

[270] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721–2726, 1999.

[271] I. Kosztin, B. Barz, and L. Janosi. Calculating potentials of mean force and diffusion coefficients from nonequilibrium process without jarzynski's equality. *J. Chem. Phys.*, 124:064106, 2006.

[272] M. B. Kubitzki and B L. de Groot. Molecular Dynamics Simulations using Temperature Enhanced Essential dynamics Replica EXchange (TEE-REX). *Biophys. J.*, 2007.

[273] L. Stryer. *Biochemistry*. W. H. Freeman and Company, 4 edition, 1995.

[274] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewskai, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. Gaussian 98, revision a.9. Technical report, Gaussian, Inc., Pittsburgh PA, 1998, 1998.

[275] W. L. Delano. *The PyMOL User's Manual*. DeLano Scientific, San Carlos, CA, USA, 2002.

[276] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[277] W. Damm, A. Frontera, J. Tirado-Rives, and W.L. Jorgensen. Opls all-atom force field for carbohydrates. *J. Comp. Chem.*, 18:1955, 1997.

[278] S. Scheuring, F. Reiss-Husson, A. Engel, J. L. Rigaud, and J. L. Ranck. High-resolution AFM topographs of rubrivivax gelatinosus light-harvesting complex lh2. *EMBO J*, 20:3029–3035, 2001.

[279] C. Möller, M. Allen, V. Elings, A. Engel, and D. J. Müller. Tapping-mode atomic force microscopy produces faithful high-resolution images of protein surfaces. *Biophys. J.*, 77:1150–1158, 1999.

[280] W. Han, S. M. Lindsay, M. Dlakic, and R. E. Harrington. Kinked dna. *Nature*, 386:563, 1997.

[281] T. E. Fisher, P. E. Marszalek, and J. M. Fernandez. Stretching single molecules into novel conformations using the atomic force microscopy. *Nat. Struct. Biol.*, 7:719–724, 2000.

Abstract (English)

 Processive enzymes are a special class of enzymes which presumably remain attached to their polymeric substrates between multiple rounds of catalysis. Due to this property, the substrate slides along the enzyme and reduces the time for the random diffusional enzyme-substrate encounters thereby increasing the efficiency of these enzymes manifold. Although structural information from many processive enzymes is available, the atomistic details of particularly the substrate sliding process, which is an inherently dynamic process, remain largely unknown. We take first steps to understand the sliding process by investigating a prototypic processive enzyme: *Streptococcus Pneumoniae* Hyaluronate lyase, a bacterial enzyme that degrades the polysaccharide substrate hyaluronan. Here we have investigated the flexibility of the enzyme as observed from essential dynamics simulations and its relation to the enzyme-substrate interactions by employing several free and enforced molecular dynamics simulations (on sub-microsecond timescale). This way we have identified a coupling between domain motions of the enzyme and the processivity or the sliding phase of the substrate. In the putative mechanism for the substrate translocation phase we observed an energy barrier along the processive direction and it is speculated that this may arise because of the reorientation of the sugar inside the cleft of the protein. This view was supported from the Force probe molecular dynamics simulations and umbrella sampling simulations that were employed to obtain a preliminary free energy profile underlying the mechanism. The observed free energy barrier is low enough to be easily crossed by thermal fluctuations, renderring essential slow collective domain rearrangements as likely rate-limiting factor for the processive cycle. The collective conformational motions of the protein along with particular interactions of individual amino acids may be involved in this translocation phase. Experimental validation along with further computational studies will be useful to understand this complex mechanism.

Keywords : *Streptococcus Pneumoniae*, bacterial hyaluronidases, hyaluronan, molecular dynamics, essential dynamics, collective motions, Force Probe Molecular Dynamics.

Abstract (Deutsch)

Prozessive Enzyme bilden eine spezielle Klasse von Enzymen, die vermientlich während vieler katalytischer Durchläufe an ihrem Substrat haften bleiben. Hierdurch gleitet das Substrat am Enzym entlang und die Zeit bis zu einem zufälligen diffusiven Aufeinandertreffen von Enzym und Substrat wird reduziert, was die Effizienz der Enzyme um ein Vielfaches erhöht. Obwohl Informationen über die Struktur vieler Prozessiver Enzyme zur Verfügung stehen, ist die Gleitphase, die inhärent dynamisch ist, weitgehend unbekannt. Wir unternehmen erste Schritte um den Gleitprozess zu verstehen, indem wir einen Prototyp für ein Prozessives Enzym untersuchen, nämlich *Streptococcus Pneumoniae* Hyaluronate lyase, ein bakterielles Enzym, das das Polysaccharidsubstrat Hyaluronsäure abbaut. Wir haben den Zusammenhang zwischen der Flexibilität des Enzyms, wie sie in "Essential Dynamics" Molekulardynamik Simulationen beobachtet wurde, mit Enzym-Substrat Interaktionen untersucht, indem wir etliche freie Simulationen und erzwungene Molekulardynamik simulationen angewandt haben. Auf diese Art haben wir eine Kopplung von Domänbewegungen des Enzyms mit der Prozessivität oder der Gleitphase des Substrates aufgezeigt. Bei dem vermuteten Mechanismus der Substrattranslokation haben wir eine Energiebarriere entlang der Prozessionsrichtung beobachtet und es ist zu vermuten, dass diese sich aus der Reorientierung des Zuckers innerhalb der Proteinspalte ergibt. Diese Sichtweise wurde unterstützt durch "Force Probe" Molekulardynamik simulationen und "Umbrella Sampling" Simulationen, die angewandt wurden um vorläufige freie Energie Profile zu erhalten, die dem Mechanismus zu Grunde liegen. Die beobachtete freie Energie Barriere ist niedrig genug um leicht durch thermische Fluktuationen überwunden zu werden, so dass essentielle langsame kollektive Reorganisationen der Domains als wahrscheinliche Raten bestimmende Faktoren für den prozessiven Zyklus in Frage kommen. Experimentelle Bestätigung zusammen mit weiteren rechnergestützten Studien wird von großem Nutzen für das Verständnis dieses komplexen Mechanismus sein.

Keywords : *Streptococcus Pneumoniae*, bakterielle Hyaluronidasen, Hyaluronsäure, Molekular dynamiksimulationen, Essential Dynamics, kollektive Bewegungen, Force Probe MD.

Lebenslauf von Harshad Joshi

Persönliche Daten

Name:	Harshad Joshi
Adresse:	Immanuel-Kant strasse 32, 37083 Göttingen
Geburtsdatum/-ort:	12. Nov, 1979, Pune, Indien
Nationalität:	Indisch

Ausbildung:

09.1992 – 06.1995	Besuch der Sou. Vimalabai Garware highschool Prüfungsfächer der Abschlussprüfung: Mathematik, Naturwissenschaften und Sozialwissenschaften. Note: First Class with Distinction.
06.1995 – 05.1997	Abitur am Fergusson College, Pune, Indien Prüfungsfächer: Physik, Mathematik, Elektronik.
06.1997 – 05.2000	Fergusson College, Pune university, Pune,Indien B.Sc.in Physik. Nebenfächer: Mathematik, Elektronik. Note: First Class with Distinction.
06.2000 – 05.2002	Pune university, Pune, Indien M.Sc. in Physik. Thema: Quantenfeldtheory. Note: First Class Thesis: *Study of effect of external perturbative electrostatic fields on polarizable molecules.* Betreuer: Dr. R. K. Pathak, Pune university, Pune, Indien.
10.2002 – 09.2003	Projektassistent in Bioinformatik. University of Helsinki, Finnland.
01.2004 –	Max Planck Institut für biophysikalische Chemie, Göttingen. Beginn der Doktorarbeit mit dem Arbeitstitel: *Functional Domain Motions and Processivity in Bacterial Hyaluronate Lyase: A molecular dynamics study.* Betreuer: Dr. Bert de Groot