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

Harshad Joshi
To my parents and teachers

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

0 Prologue .......................................................... xi
  0.1 Scope of the thesis ........................................... xiii
  0.2 Outline of the thesis ......................................... xvi

1 Theory and Methods .............................................. 1
  1.1 Molecular Dynamics simulations - Principle ................. 3
    1.1.1 From Schrödinger equation to Molecular Dynamics .... 4
  1.2 MD simulations in practice .................................. 9
    1.2.1 Integration method and time step ....................... 9
    1.2.2 Solvent environment ..................................... 10
    1.2.3 System boundaries ....................................... 11
    1.2.4 Temperature and pressure coupling ..................... 12
    1.2.5 Improving efficiency .................................... 14
    1.2.6 Minimisation and equilibration ........................ 16
  1.3 Analysis methods ............................................ 17
    1.3.1 Principal Component Analysis .......................... 17

2 Processivity and Hyaluronidases ................................. 21
  2.1 Processive Enzymes .......................................... 23
  2.2 Hyaluronidases ............................................... 25
    2.2.1 History and physiological importance .................. 26
    2.2.2 Classification of Hyaluronidases and *Spn*.Hyaluronate Lyase .... 27
  2.3 *Streptococcus Pneumoniae* Hyaluronidase (*Spn*.Hyal) .... 28
    2.3.1 Choice of *Spn*.Hyal for processivity study .......... 29
    2.3.2 Structure of *Spn*.Hyal and its catalytic mechanism ... 29
    2.3.3 Previous studies on the processive mechanism of *Spn*.Hyals ... 31
3 Hyaluronan: Structure and dynamics

3.1 Hyaluronan (HA): Introduction

3.1.1 Chemical structure and polymer properties

3.1.2 Solution structure and aggregation properties

3.1.3 Occurrence and physiological importance

3.1.4 Medical applications

3.2 Force field choice

3.2.1 Simulation setup

3.2.2 Validation and choice of OPLS-SEI* force field

3.2.3 Flexibility of HA in the presence of the Spn. Hyal

3.3 Conclusions

4 Flexibility of bacterial Hyals: Part I

4.1 Conformational motions of proteins

4.2 Crystallisation of new Spn. Hyal structures

4.2.1 Crystallisation and diffraction data collection

4.2.2 Structure solution and refinement

4.3 Simulation details

4.4 Results

4.4.1 Description of the new pneumococcal Hyal structures

4.4.2 Bacterial Hyals are flexible

4.5 Discussion

4.6 Conclusions

5 Flexibility of bacterial Hyals: Part II

5.1 Flexibility of Spn. Hyaluronate Lyases

5.2 Simulation details

5.3 Results and Discussions

5.3.1 Equilibration and the flexibility of the Spn. Hyals

5.3.2 PCA and the domain motions in Spa. Hyals

5.3.3 Flexibility of the HA inside the cleft

5.3.4 Processive motion of the sugar

5.3.5 Role of protein dynamics in substrate binding
## List of Figures

1. Schematic representation of simulation method dependence on the size of the system and the timespan of the process of interest. .......... xiii
2. A typical Spn.Hyal-HA simulations setup. ............................. xv

1.1 Illustration of the force field terms. ........................................ 7
1.2 Typical molecular dynamics simulation system with explicit solvent box. 13
1.3 A schematic representation of Principal Component Analysis (PCA) for the MD-trajectory. .................................................... 18

2.1 Representative structures for each class of processive enzymes. ........... 25
2.2 Schematic diagram of the virulence factors of Strep. Pneumoniae. .......... 26
2.3 Structural details of Spn.Hyal. with the electrostatic potential distribution in the catalytic cleft. ............................................. 30
2.4 Details of the catalytic mechanism in Spn.Hyal-HA system. ................. 32
2.5 The three largest domain motions of Spn.Hyal as obtained by CONCOORD. 33

3.1 Chemical structure of Hyaluronic acid. ..................................... 37
3.2 Solution structure of Hyaluronic acid I — Hydrogen bonding. ............... 39
3.3 Solution structure of Hyaluronic acid II — Aggregation properties. ....... 40
3.4 2D projections of simulations with different force fields on the PCA eigenvector-set of experimental structures. .......................... 44
3.5 Dihedral distribution along the glycosidic linkages of HA6 polymer for the simulations with different force fields. ......................... 45
3.6 Difference in dihedral distribution along the glycosidic linkages of HA6 polymer in the presence or absence of the Spn.Hyal. ............... 46

4.1 A schematic representation of the relationship of the flexibility of the protein and the energy landscape explored by it. ....................... 51
4.2 Difference in the extent of the opening/closing of the catalytic cleft for different crystal structures discussed in the chapter. .................................. 56

4.3 MD simulation trajectories show that the flexibility of the Spn.Hyal as affected by the presence of the HA hexasaccharide inside the cleft. Simulation with holo structure remains ‘locked’ while the cleft opens significantly in the apo-simulation. ................................................................. 57

5.1 Root mean square fluctuations (RMSF) per residue for the simulations show the three loop regions around the cleft of the Spn.Hyal are highly flexible. ................................................................. 68

5.2 2-D Projections of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. These projections show the flexibility of the protein along the three domain motions observed in Chapter 4 .................................. 70

5.3 Time evolution of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. ........................................ 71

5.4 Motion of the sugar inside the cleft measured as the distance from the catalytic cleft. ................................................................. 76

5.5 Flexibility of the sugar inside the cleft measured as the root mean square deviation (RMSD) of the first four rings of HA with respect to those in the crystal structure 1LOH.pdb ................................................................. 77

5.6 Motion of the sugar in the processive direction inside the cleft as depicted by the PCA performed on the sugar. ................................................................. 78

5.7 Short range interactions of the sugar with the positive and the aromatic region in the cleft. ................................................................. 82

5.8 Spn.Hyal-HA interface region showing important residues that are in contact with the HA in the crystal structure 1LOH.pdb ................................................................. 84

5.9 Short range interactions of the individual sugar rings with the specific residues that are known from the mutation experiments or are thought to play a key role in the processivity mechanism. ................................................................. 86

5.10 General regions around the catalytic cleft. ................................................................. 88

5.11 Short range interactions of the last disaccharide units (at non-reducing end) of the HA with the entry region of the cleft. ................................................................. 89

6.1 Schematic representation of the energy landscapes explored by the system at different temperatures. ................................................................. 94
6.2 2-D & 1-D Projections of the simulation trajectories together with the x-ray structures projected on the essential subspace obtained from the PCA. The increased temperature of the sugar affected the flexibility of the protein indicating strong coupling of the protein dynamics to the sugar mobility. 96

6.3 Flexibility of the sugar inside the cleft for the high-temperature simulations as measured from the RMSD and the PCA of the first four rings of the sugar. 98

6.4 A typical Force Probe Molecular Dynamics (FPMD) simulation for Spn.Hyal-HA system. 100

6.5 Schematic representation of Essential Dynamics technique. 101

6.6 FPMD simulation set up for the Spn.Hyal-HA system. 102

6.7 Effect of different force constants and pulling velocities on the rupture forces. 105

6.8 Scheme of the FPMD simulations setup to investigate the protein flexibility during the sugar translocation. 106

6.9 Relationship between the protein flexibility and the sugar translocation through the cleft. 107

6.10 Rupture forces for different FPMD simulations and a probable explanation for the easier pulling in the opposite direction. 108

6.11 Short range interactions of the sugar with the cleft of the Spn.Hyal in different EDSAM simulations. 111

6.12 Free energy estimates for the sugar translocation along the putative reaction pathway from different enforced simulations. 114

B.1 Schematic representation of HA disaccharide unit with the atom numbers labeled for the topology. 129

C.1 A schematic representation of a typical AFM experiment. 133
List of Tables

1.1 Typical vibration periods in molecules and hydrogen-bonded liquid. Compare $\hbar/kT = 170\,fs$ at 300 K .............................................. 6

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. ........................................... 42

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 computationally removed from the crystal structure. ........................................ 54

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-PEGMME refer to new structures solved while SagHyal refer to structures from different bacterial species. ............................................. 54

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 $C^\alpha$ 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. .................. 59

5.1 List of simulation systems with the simulation time in ns (the 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. ............................................. 66
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 $C^\alpha$ 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.

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

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 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 $C^\alpha$ 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.

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

B.2 Sets of Tosional Angle parameters for Hexapyranoses for the OPLS-AA and OPLS-SEI force fields corresponding to the Potential Energy Function according to Kony et.al.

B.3 List of torsional angles that were modified according to new OPLS-SEI force fields.
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:\textsuperscript{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 these ‘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 or modern techniques like x-ray free electron lasers (xFEL), cryoelectron tomography 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, single molecule atomic-force microscopy (AFM) combined with either optical or magnetic tweezers, NMR relaxation measurements. 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 are 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 chosen carefully, it has shown to be quite powerful to investigate many different processes. The examples include the reversible folding of peptides, or large scale conformational changes in ATPase or chaperones. Such dynamical structural changes are results of the exploration of these molecules in their high-dimensional free energy landscape, which are often required to fulfil their diverse functions. 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
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 so 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 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. 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 λ-exonuclease which has a closed toroidal structure that degrades double-stranded DNA (dsDNA) into a single-stranded DNA (ssDNA) or a recently solved phosphomannomutase/phosphoglucomutase (PMM/PGM), 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. 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 an energetic or bioengineering point of view, for example to engineer a processive enzyme mimic.

**Molecular mechanism of sugar degradation by hyaluronidases**

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

Recently, high-resolution structures of hyaluronate lyase from two different species (*Streptococcus Pneumoniae* and *Streptococcus Agalactiae*) were solved, both of the apo form of the enzymes, as well as in complex with substrate and product. 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
0.1 Scope of the thesis

A typical *Spn.*Hyal-HA system 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 is having 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.

mammalian Hyals) even in complex with HA tetrasaccharide (pdb code: 1FCV (complex with HA) and native enzyme 1FCQ, 1FCU). As these studies 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 questions for each of the studies is addressed in details in the introductory section of the corresponding chapters (Chapter 3 through Chapter 6). The thesis is structured as follows:

Chapter 1 — Introduction

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. 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 so far categorised 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 a 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 of role of conformational change of the protein to be 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 are sketched in this chapter.
Chapter 3 — Hyaluronan oligomers

The polysaccharide hyaluronan that is processively degraded by *Spn*.Hyal has been widely studied itself 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 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 *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 \(\sim 100\) 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

\(^*\)when Kramer’s rate theory is applied
the 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. 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.

Finally in Chapter 7, we summarise all the results from these studies. The general mechanism of 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) method, that are used throughout the present work are described in Sec. 1.3.
When it all began...

Newton’s *Principia Mathematica*\textsuperscript{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 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”\textsuperscript{35} when classical Newtonian mechanics was found to be inadequate to explain the phenomena on the atomic scale, and the theory of ‘Quantum Mechanics’\textsuperscript{36,37} was born. Despite the philosophical questions of interpretation\textsuperscript{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 \textit{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.\textsuperscript{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 manageable 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, \textit{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\textsuperscript{40–43} and many textbooks.\textsuperscript{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