Western University

Scholarship@Western

2011

MOLECULAR DYNAMICS ANALYSIS OF THE BEHAVIOUR OF PROLYL OLIGOPEPTIDASE (POP) IN THE PRESENCE OF Z-PRO-PROLINAL INHIBITOR

Sepideh Soltani

MOLECULAR DYNAMICS ANALYSIS OF THE BEHAVIOUR OF
PROLYL OLIGOPEPTIDASE (POP) IN THE PRESENCE OF
Z-PRO-PROLINAL INHIBITOR
(Spine title: MD Analysis of Behaviour of POP in Presence of ZPP Inhibitor)
(Thesis format: Monograph)

by

Sepideh Soltani

Graduate Program in Applied Mathematics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Sepideh Soltani 2011

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Examiners:

. . . .
Dr. A. Buchel

Supervisor:

. . . .
Dr. A. Metzler

. . . .
Prof. M. Karttunen

. . . .
Dr. G. Hunter,

The thesis by

Sepideh Soltani

entitled:

Molecular Dynamics Analysis of the behaviour of Prolyl Oligopeptidase (POP) in the presence of Z-Pro-Prolinal inhibitor

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

. . . .
Date   Chair of the Thesis Examination Board

Abstract

Prolyl Oligopeptidase (POP), a member of the prolyl endopeptidase family, has a role in several neurological disorders. Its primary function is to cleave an oligopeptidase, including neuroactive peptides. On the other hand POP with a Z-Pro-prolinal (ZPP) inhibitor may revert memory loss from neurological disorders, amnesic agents and aging. Here, the crystal structure of POP protein with ZPP inhibitor (Protein Data Bank PDB) and without ZPP inhibitor is studied using classical molecular dynamics simulations and the POP-ZPP complex behaviour is compared with pure POP. The basic analysis of the structures, included measuring radius of gyration and root mean square deviation which proved that POP structure with non-bonded ZPP and without ZPP are stable and maintain their structure over the entire simulation time. Moreover, principal component analysis (PCA) is used to analyze the motions of the structures by extracting the normal modes of motions in POP with and without presence of ZPP inhibitor.

Acknowlegements

First and foremost, I would like to thank my supervisor, Dr. Mikko Karttunen, who not only came up with the project, but also kept pushing me throughout the years. This is especially true for the last 2 years during which, without his constant support, it is unlikely this project would have been completed. I would like to acknowledge my friends: Amir Mohsen Pourmousa, Sussana Hug and Elio Cino. I am thankful for the faculty, staff and fellow students who made my stay at the university enjoyable. Last, but not least, I would like to extend my gratitude to my husband Ashkan Rasouli and parents, who supported me throughout this adventure and put up with me when I was being difficult, which apparently occurred more often than I thought.

Contents

List of Figures

List of Appendices

Chapter 1

Introduction

The building blocks of all living organisms are cells [4]. Cells can be divided into two large groups: prokaryotic and eukaryotic cells. Eukaryotes have a nucleus and prokaryotes don't. [120].

Eukaryote has a plasma membrane that separates the interior of the cell from the surrounding. Everything inside of the cell is called cytoplasm. The cytoplasm contains all the structures needed to keep the cell alive and functioning including the cell nucleus. Most of the dry mass of a cell is composed of proteins. To analyze the electrical or biochemical activity of cells, information can be obtained by studying the proteins. In addition to being the building blocks of the cells, the proteins carry out almost all of the cell functions. For example, enzymes (which are a type of protein) provide sophisticated molecular surfaces that are required for many chemicals reactions inside the cells.

Proteins inserted in the plasma membrane form channels and pumps that control the entry and exit of small molecules into or out of the cell. Some of the proteins carry messages from cell to cell, while others transmit signals from the plasma membrane into the nucleus. Some proteins act as very small molecular machines with moving parts to propel molecules inside the cell (e.g. *kinesin*), untangle supercoiled DNA molecules (e.g. *topoisomerase*) etc. [2]. Other specialized proteins perform as antibodies, hormones, toxins, anti-freeze molecules, elastic fibres, ropes or sources of luminescence [2].

Therefore, in order to understand the behaviour of cell and the human body, proteins must be studied first. Proteins are divided into different categories based on their functionality. The most important protein groups are enzymes (catalysts of chemical reactions), structural proteins (supporters of the cell and tissue structure), transport proteins (carriers of small molecules or ions), motor proteins (movement creators), storage proteins (storing small molecules or ions), signalling protein (carriers of signals between cells), receptor proteins (signal detectors and transmitters working through the cell), gene regulatory proteins (turning genes on and off by binding with them) and special purpose proteins.
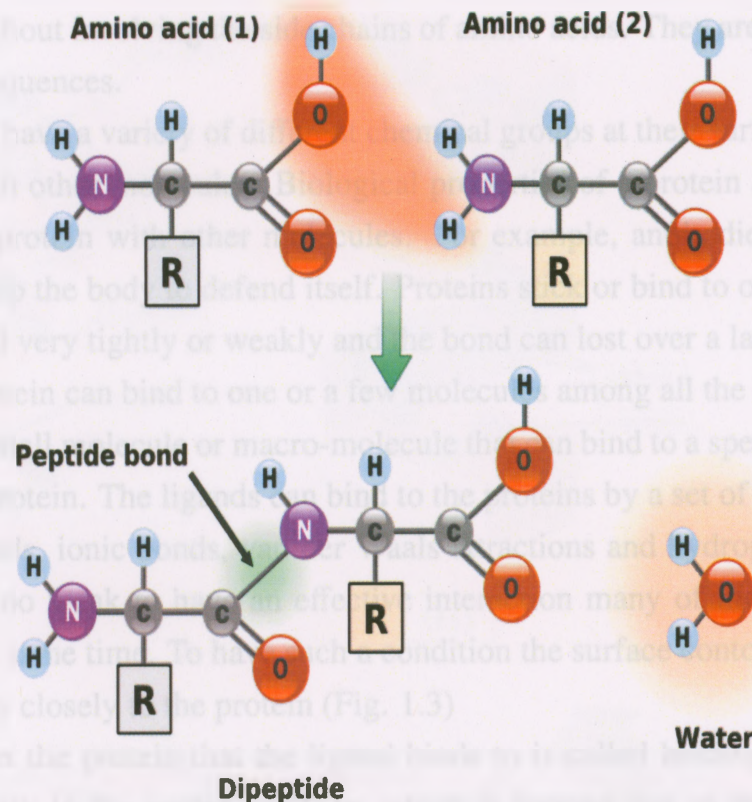
Figure 1.1: A peptide bond. This covalent bond forms when the carbon atom from the carboxyl group of the amino acid shares electrons with the nitrogen atom (blue) from the amino group of the second amino acid. As indicated, a molecule of water is lost in this condensation reaction. The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia. [112].

Proteins have very complex structures, which allow them to function for different purposes. Protein molecules are made up of a long chain of amino acids which are linked to their neighbors by covalent peptide bonds (Fig. 1.1). The repeating sequence of atoms along the chain is called the polypeptide back bone. Each type of protein has a different sequence of amino acids (differs in their side-chains), accounting for the diversity of proteins. Most of the covalent bonds within the chain let the joining atoms rotating freely about the bond. Therefore, the polypeptide back bone can fold up in different ways where the system has the lowest energy. Each of the folded chains is constrained through many different non-covalent bonds, which are formed by atoms in the polypeptide backbone and by atoms in the amino acid side chains [2]. These weak bonds are hydrogen bonds, ionic bonds and van der Waals attractions described in chapter three.

In most of the proteins two regular patterns (secondary structure) are most likely to be present. The folding patterns are α helix and β sheets (Fig. 1.2). These two patterns are very common because they result from hydrogen bonding between N-H and C=O groups in the polypeptide

back bone without involving the side chains of amino acids. They are formed by many different amino acid sequences.

The proteins have a variety of different chemical groups at their surface, which define the way they react with other molecules. Biological properties of a protein are defined by the interactions of that protein with other molecules. For example, antibodies can attach to viruses or bacteria to help the body to defend itself. Proteins stick or bind to other molecules differently. They can bind very tightly or weakly and the bond can lost over a larger or shorter time. However, each protein can bind to one or a few molecules among all the molecules that come close to it. A ion, small molecule or macro-molecule that can bind to a specific protein is called a ligand for that protein. The ligands can bind to the proteins by a set of weak non-covalent bonds: hydrogen bonds, ionic bonds, van der Waals attractions and hydrophobic interactions. Since one bond is too weak to have an effective interaction many of those weak bonds should be formed at the same time. To have such a condition the surface contours of the ligand molecule should fit very closely to the protein (Fig. 1.3)

The region in the protein that the ligand binds to is called binding site. The binding site is usually a cavity in the protein surface, which is formed due to the specific arrangement of amino acids (tertiary structure of a protein). The amino acids in the binding site can belong to different parts of the polypeptide chain that have come close together when the protein folds. Different regions of the protein surface can provide binding sites for different ligands.

In chapter two, further biophysical background on amino acids and prolyl oligopeptidases is presented. Prolyl Oligopeptidase family is a group of enzymes that contains four enzymes: prolyl oligopeptidase (POP), oligopeptidase B, dipeptidyl-peptidase IV and acylaminoacyl peptidase. POP has a role in cell degeneration and apoptosis [83] which lead to characteristic cell changes and death but specific inhibitors of this peptidase (Z-pro-prolinal) reverts memory loss produced by neurological disorders, amnesic agents and ageing [168]. Therefore we studied POP with ZPP in my work.

Chapter three, provides a brief background on the methods. To study the dynamic properties of POP, we used molecular dynamics. In this chapter MD's subdivisions are explained while focusing on classical molecular dynamics. Moreover, different kinds of potentials and force fields are discussed and the suitable ones for the purpose of the present work are selected. The thermostats and their specific behaviour are discussed and the proper one is selected.

In chapter four, the procedure to setup the classical molecular dynamics simulations, details of the simulation and the analysis are explained.

Finally, in chapter five the results of the simulations are presented. The dynamic properties of the POP with and without a ligand inhibitor (ZPP) are compared.
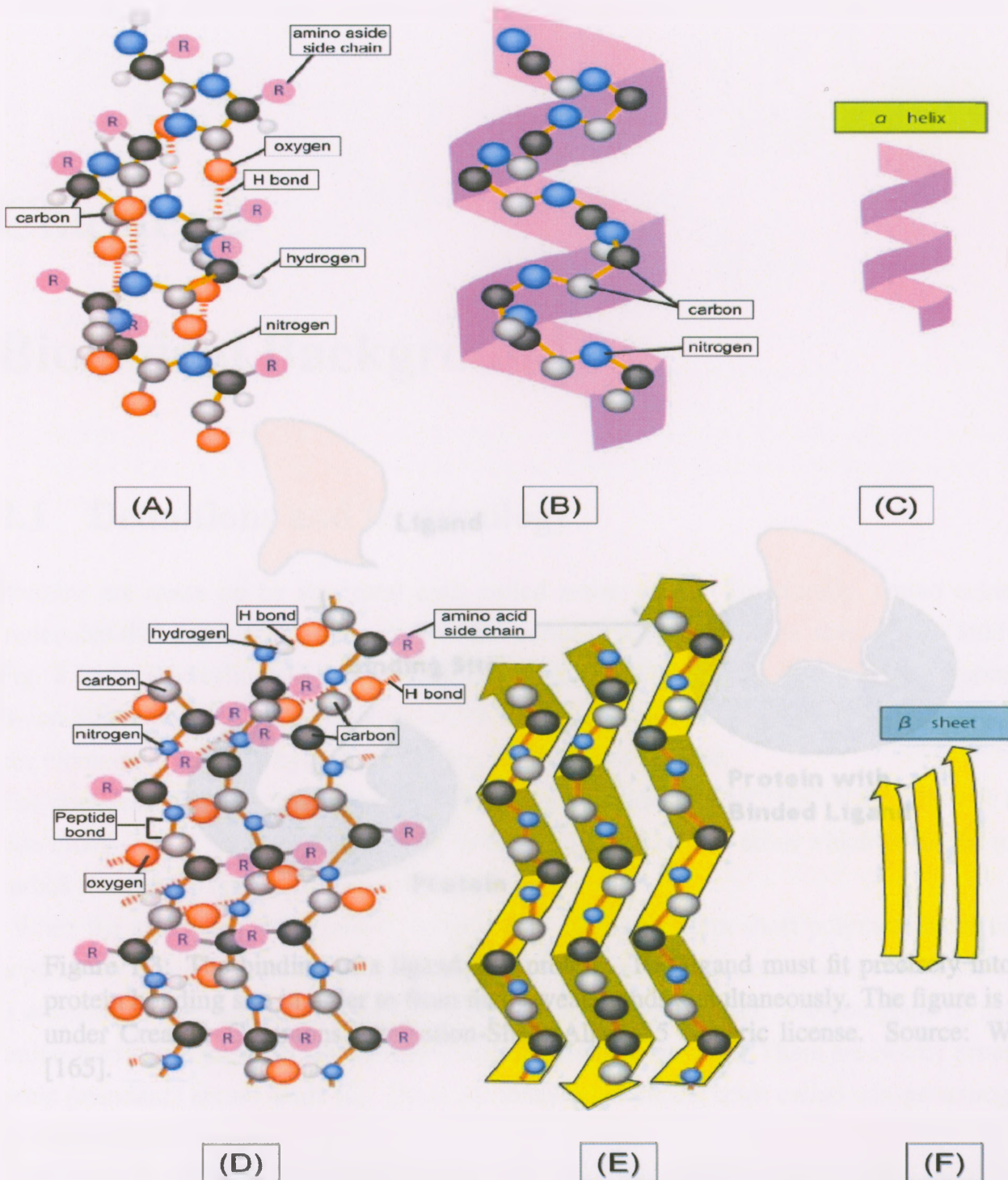
Figure 1.2: The regular conformation of the polypeptide backbone observed in the α helix and β sheet. (A),(B) and (C) display the α helix, in which the N-H of the peptide bond is hydrogen-bonded to the C=O of a neighboring peptide bond located four amino acids away in the same chain. (D), (E) and (F) display a β sheet; in this secondary structure, adjacent peptide chains run in opposite (anti parallel) directions. The individual polypeptide chains (strands) in a β sheet are held together by hydrogen-bonding between peptide bonds in different strands, and the amino acid chains in each strand alternately project above and below the plane of the sheet. (A) and (D) show all the atoms in the polypeptide backbone, but the amino acid side chains are truncated and denoted by R. In contrast, (B) and (E) show the backbone atoms only, while (C) and (F) display the shorthand symbols that are used to represent the α helix and β sheet in ribbon drawing of proteins. The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Pukiwiki [84].

Chapter 2

Biological Background

2.1 Definition and terminology
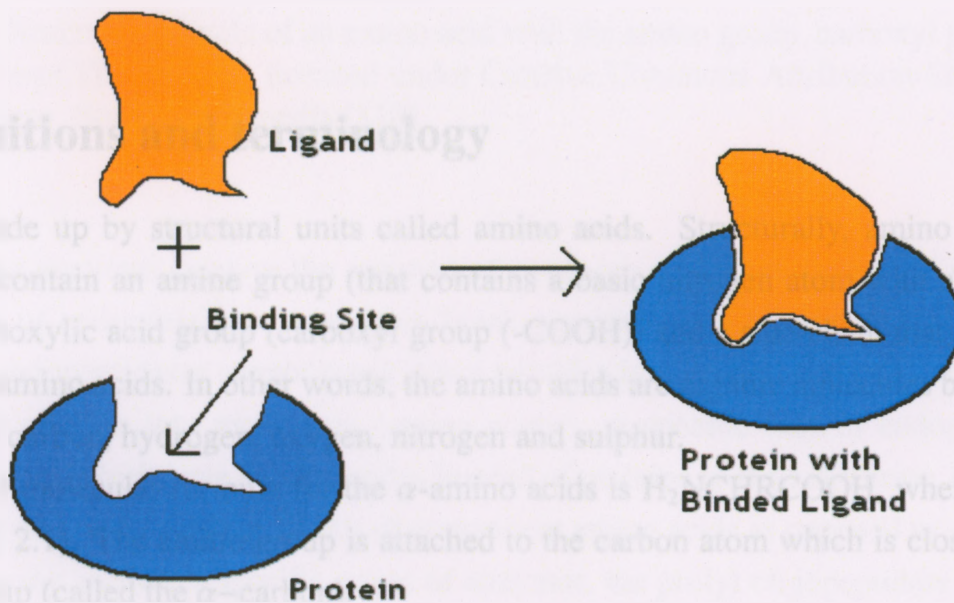
Figure 1.3: The binding of a ligand to a protein. The ligand must fit precisely into the the protein bonding site in order to from many weak bonds simultaneously. The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia [165].

Chapter 2

Biological Background

2.1 Definitions and terminology

Proteins are made up by structural units called amino acids. Structurally, amino acids are molecules that contain an amine group (that contains a basic nitrogen atom with a lone pair Fig. 2.1), a carboxylic acid group (carboxyl group (-COOH)) and a side-chain that varies between different amino acids. In other words, the amino acids are organic molecules built of the key elements of carbon, hydrogen, oxygen, nitrogen and sulphur.

The chemical or molecular formula for the α-amino acids is $H_2NCHRCOOH$, where R is an side-chain (Fig. 2.1). The amino group is attached to the carbon atom which is closest to the carboxylate group (called the α-carbon).

When the amino acids are joined together they can either form short polymer chains called peptides or longer chains called polypeptides or proteins. Each amino acid within the chain is attached to two neighboring amino acids except ones at the end. Proteinogenic or natural amino acids are a group of amino acids found naturally in proteins. There are twenty proteinogenic (standard) amino acids [2]. There are many other amino acids called non-proteinogenic or non-standard.

Within the 22 standard amino acids, 20 are geneticaly encoded (Fig. 2.2), Cystine and hydroxyproline are uncommon amino acid which are not shown in the figure. Cystine is a dimeric amino acid formed by the oxidation of two cysteine residues that covalently link to make a disulfide bond. Hydroxyproline differs from proline by the presence of a hydroxyl (OH) group attached to the C atom. Proteins that increase the rate of chemical reactions (working as catalyzers) are called enzymes. In enzymatic reactions, the molecules at the beginning of the process (substrates) are converted into different molecules (products). Almost all chemical reactions in a biological cell are accelerated by enzymes.

Enzymes that cut peptide bonds are called proteinases or peptidases, e.g. serine peptidases.
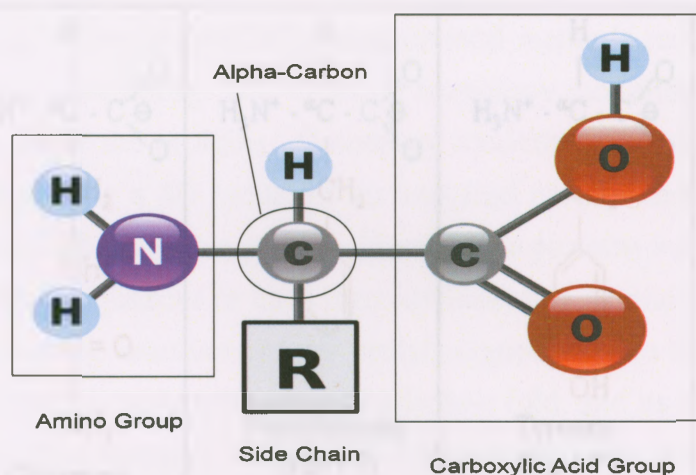
Figure 2.1: Structural formula of an amino acid with the amino group, carboxyl group and side chain (R group).The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia [118]

Exopeptidase enzymes catalyse the removal of an amino acid from the end of a protein. Endopeptidase or endoproteinase peptidases break peptide bonds of nonterminal amino acids. Therefore, endopeptidases cannot break down peptides into monomers but exopeptidases can break down proteins into monomers (amino acids). A specific case of endopeptidases are oligopeptidases, which have oligopeptides instead of proteins as their substrate. Oligopeptides are peptides of a small number (3-40) of component amino acids as opposed to a polypeptide. In the following section, a specific group of enzymes, the prolyl oligopeptidase family, is explained in order to create a better understanding of the subject of this thesis.

2.2 The Prolyl Oligopeptidases (POPs)

The prolyl oligopeptidase family (POPs) are enzymes that belongs to a new class of serine peptidases and cannot hydrolyze peptides containing more than about 30 residues. POPs cleave peptide bonds at the carboxy group of proline residues. POP was first discribed in the 1991. Rennex *et al.* analyzed the amino acid sequence of a prolyl endopeptidase (prototype prolyl oligopeptidase) from pig brain [137] (Fig. 2.3). Since the mentioned prolyl endopeptidase hydrolyses only oligopeptides (not proteins), in contrast with the proly endopeptidases from various species of bacteria that can degrade immunoglobulin A, it was renamed to prolyl oligopeptidase (prototype prolyl oligopeptidase) [137].Then dipeptidyl peptidase IV, acylaminoacyl peptidase and oligopeptidase B were included to the prototype prolyl oligopeptidase [137, 132] and made this POPs family (Fig. 2.3).

Arginine
(Arg / R)

Glutamine
(Gln / Q)

Phenylalanine
(Phe / F)

Tyrosine
(Tyr / Y)

Tryptophan
(Trp, W)

Lysine
(Lys / L)

Glycine
(Gly / G)

Alanine
(Ala / A)

Histidine
(His / H)

Serine
(Ser / S)

Proline
(Pro / P)

Glutamic Acid
(Glu / E)

Aspartic Acid
(Asp / D)

Threonine
(Thr / T)

Cysteine
(Cys / C)

Methionine
(Met / M)

Leucine
(Leu / L)

Asparagine
(Asn / N)

Isoleucine
(Ile / I)

Valine
(Val / V)
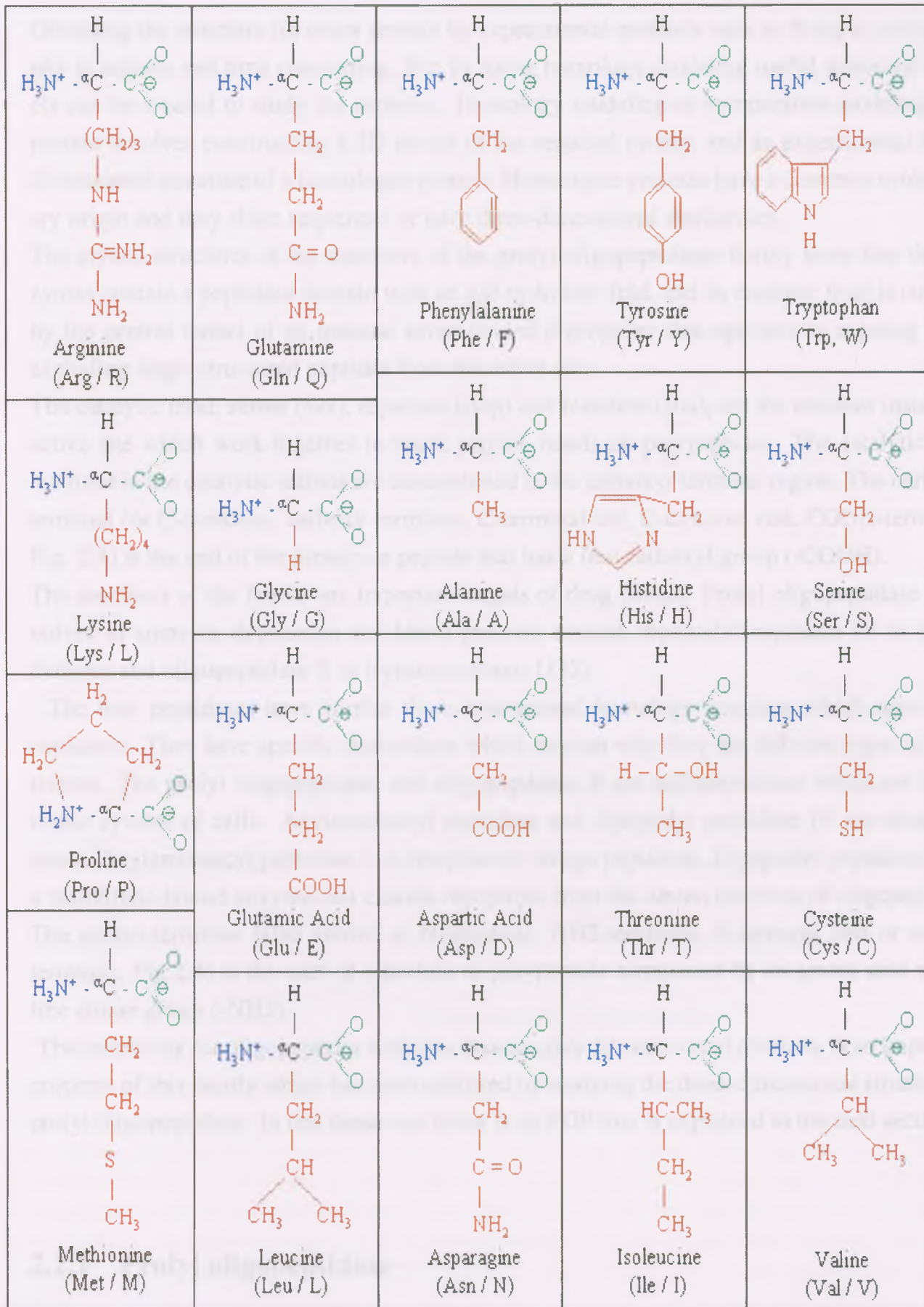
Figure 2.2: Structure of the 20 amino acids that are encoded which these amino acids found naturally in proteins.The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia [10].

Obtaining the structure for every protein by experimental methods such as X-ray crystallography is tedious and time consuming. But by using homology modeling useful structural models can be created to study the proteins. Homology modeling or comparative modeling of a protein involves constructing a 3D model of the required protein and an experimental three-dimensional structure of a homologue protein. Homologue proteins have a common evolutionary origin and they share sequences or have three-dimensional similarities.

The crystal structures of the members of the prolyl oligopeptidases family show that the enzymes contain a peptidase domain with an α/β hydrolase fold, and its catalytic triad is covered by the central tunnel of an unusual seven-bladed β-propeller that operates as a gating filter, excluding large, structured peptides from the active site.

The catalytic triad, serine (Ser), aspartate (Asp) and histidine (His), are the residues inside the active site which work together to break peptide bonds on polypeptides. The catalytic triad involved in the catalytic actions are concentrated in the carboxyl terminal region. The carboxyl terminal (or C-terminus, carboxy-terminus, C-terminal tail, C-terminal end, COOH-terminus, Fig. 2.4) is the end of the protein or peptide that has a free carboxyl group (-COOH).

The members of the family are important targets of drug design. Prolyl oligopeptidase is involved in amnesia, depression and blood pressure control, dipeptidyl peptidase IV in type 2 diabetes and oligopeptidase B in trypanosomiasis [132].

The four peptidases have similar three-dimensional homology structure which have been explained. They have specific distinctions which explain why they are different types of peptidases. The prolyl oligopeptidase and oligopeptidase B are endopeptidases which are found in the cytosol of cells. Acylaminoacyl peptidase and dipeptidyl peptidase IV are exopeptidases. Acylaminoacyl peptidase is a cytoplasmic omega peptidase. Dipeptidyl-peptidase IV is a membrane-bound enzyme that cleaves dipeptides from the amino terminus of oligopeptides. The amino terminus (also known as N-terminus, NH2-terminus, N-terminal end or amine-terminus, Fig.2.4) is the start of a protein or polypeptide terminated by an amino acid with a free amine group (-NH2).

The selectivity for oligopeptides with less than roughly 30 amino acid residues, is an important property of this family which has been analyzed by studying the three-dimensional structure of prolyl oligopeptidase. In this thesis our focus is on POP that is explained in the next section.

2.2.1 Prolyl oligopeptidase

POP was discovered in the human uterus as an oxytocin-degrading enzyme [101, 163] (oxytocin is a mammalian hormone acts primarily as a neuromodulator). This peptidase was orig-

```
1        10        20        30        40        50
MLSFQYPDVYRDETAIQDYHGHKVCDPYAWLEDPDSEQTKAFVEAQNKITVPFLE
              V         I
                β1*      β2*                    α1*

     60        70        80        90        100       110
QCPIRGLYKERMTELYDYPKYSCHFKKGKRYFYFYNTGLQNQRVLYVQ■SLEGEA
        α2*              β1/1     β1/2          β1/3

     120       130       140       150       160
RVFLDPNILSDDGTVALRGYAFSEDGEYFAYGLSASGSDWVTIKFMKV■GAKELP
β1/4              β2/1          β2/2          β2/3

     170       180       190       200       210       220
DVLERVKFSCMAWTHDGKGMFYNAYPQQDGKSDGTETSTNLHQKLYYHVLGT■QS
                                S
β2/4       β3/1        β3/2                      β3/3

     230       240       250       260       270
EDILCAEFPDEPKWMGGAELSDDGRYVLLSIREGCDPVNRLWYC■LQQESNGITG
                                                    S  A
 β3/4        β4/1         β4/2          β4/3

     280       290       300       310       320       330
ILKWVKLIDNFEGEYDYVTNEGTVFTFKTNRHSPNYRLINI■FTDPEESKWKVLV
                                Q          V        W
    β4/4      β5/1       β5/2          β5/3            β5/4

     340       350       360       370       380
PEHEKDVLEWVACVRSNFLVLCYLHDVKNTLQLH■LATGALLKIFPLEVGSVVGY
        I                        I        T      T    D      I
    β6/1        β6/2         β6/3       β6/4        β7/1

     390       400       410       420       430       440
SGQKKDTEIFYQFTSFLSPGIIYHC■LTKEELEPRVFREVTVKGIDASDYQTVQI
     β7/2          β7/3          β7/4          β1

     450       460       470       480       490
FYPSKDGTKIPMFIVHKKGIKLDGSHPAFLYGYGGFNISITPNYSVSRLIFVRHM
  β4         β2                β3              αA

     500       510       520       530       540       550
GGVLAVANIRGGGEYGETWHKGGILANKQNCFDDFQCAAEYLIKEGYTSPKRLTI
   I
   β4         αB'           αB                      β5

     560       570       580       590       600
NGG■NGGLLVATCANQRPDLFGCVIAQVGVMDMLKFHKYTIGHAWTTDYGCSDSK
                  A
        αC                β6                  αD'

     610       620       630       640       650       660
QHFEWLIKYSPLHNVKLPEADDIQYPSMLLLTADH■DRVVPLHSLKFIATLQYIV
       V
   αD                        β7                    αE

     670       680       690       700       710
GRSRKQNNPLLIHVDTKAG■GAGKPTAKVIEEVSDMFAFIARCLNIDWIP
      S                                        V
      β8                        αF
```

Figure 2.3: Amino Acid Sequence and Sec- ondary Structure of Prolyl Oligopeptidase The sequence from porcine brain (Rennex et al., 1991) is shown in the top row. Residues different in the human lymphocyte enzyme (Vanhoof et al., 1994) are shown in the second row. The third strand within the β sheets of the propeller domain always terminates with an aspartate residue (boxed). Members of the catalytic triad (Ser554, His680, and Asp641) are also boxed. The secondary structure as- signment is made according to DSSP (Kabsch and Sander, 1983). The figure was produced with ALSCRIPT (Barton, 1993). From [137] with permission. ©: Elsevier
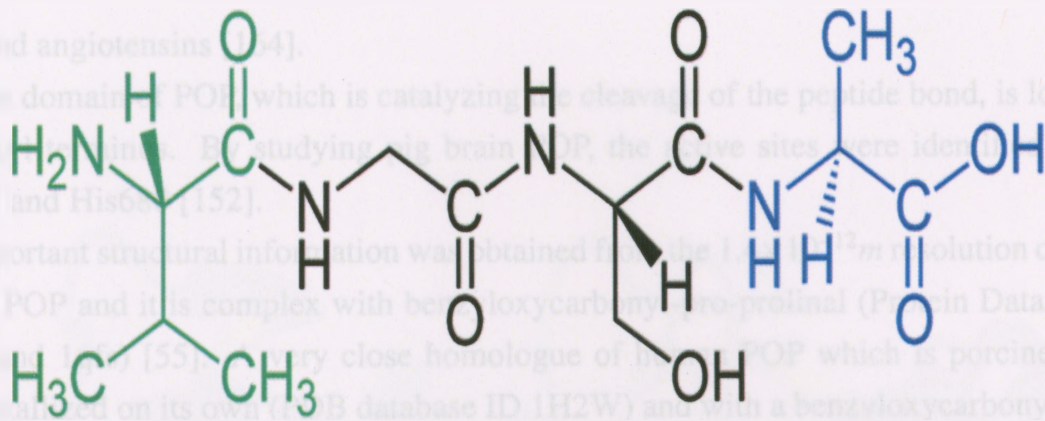
Figure 2.4: A tetrapeptide (example: Val-Gly-Ser-Ala) with green highlighted N-terminal α-amino acid (example: L-valine) and blue marked C-terminal α-amino acid (example: L-alanine). Amino terminus is the start of a protein or polypeptide which is terminated by an amino acid with a free amine group (-NH2) and The carboxyl terminal is the end of the amino acid chain that has a free carboxyl group (-COOH). The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia [92].

inally named post-proline cleaving enzyme because it hydrolyzes the peptide bond on the carboxyl side of proline residues. Later, its another name, prolyl endopeptidase, was recommended with respect to the enzyme nomenclature. Finally, it was named prolyl oligopeptidase to emphasize the special characteristics of the enzyme that are explained next [132].

POP is present in most organisms and tissues. It also has been cloned from different sources such as porcine brain [136, 139], *Flavobacterium meningosepticum* [166, 29, 44], human lymphocytes [158, 147], mouse brain [82], bovine brain [167], *Sarcophaga peregrina* [125], *Aeromonas hydrophila* [94] and *Pyrococcus furiosus* [140, 65]. The mouse prolyl oligopeptidase gene has been extensively studied and the structure and localization of it are resolved [96]. A decrease in serum POP activity has been observed in patients suffering from different stages of depression [111].

By studying inhibitors of POP in more details about physiological effects of this enzyme can has been obtained. POP has been proven to have negative effects on memory loss that is caused by neurological disorders, amnesic agents and aging [168, 132] but inhibiting POP has shown positive reactions in the mouse model of accelerated senescence by suppressing the formation of amyloid β-peptide in neuroblastoma (a type of nerve cancer) cells [146] and is thus preventing amyloid-like depositions. However, specific inhibitors do not affect the formation and degradation of β−amyloid peptides and β-amyloid precursor protein related to Alzheimer disease [131]. This behaviour can be explained by the fact that the POP does not hydrolyze large peptides and proteins. POP also has a role in the blood pressure regulation by participating in the renin-angiotensin system which regulates the blood pressure through metabolism of

bradykinin and angiotensins [164].

The peptidase domain of POP, which is catalyzing the cleavage of the peptide bond, is located at the carboxyl terminus. By studying pig brain POP, the active sites were identified to be Ser554 [139] and His680 [152].

The most important structural information was obtained from the $1.4 \times 10^{-12} m$ resolution crystal structures of POP and it is complex with benzyloxycarbonyl-pro-prolinal (Protein Data Bank codes 1qfm and 1qfs) [55]. A very close homologue of human POP which is porcine POP has been crystallized on its own (PDB database ID 1H2W) and with a benzyloxycarbonyl-pro-prolinal (also called Z-pro-prolinal (ZPP)), (Fig. 2.6) inhibitor (PDB database code 1QFS) covalently bound to the serine 554 residue through a hemiacetal bond [55]. ZPP inhibitor has a hydrophobic head which sterically blocks the active site and an aldehyde tail that forms a reversible covalent hemiacetal bond with the SER555 residue of the catalytic domain (Fig. 2.5). In this thesis we used the structure of the complex of porcine POP with ZPP (1QFS).

POP has two domains: the catalytic and propeller domain. The peptidase or catalytic domain is built up of residues $1 - 72$ and $428 - 710$. The residues $73 - 427$ build the propeller domain. The peptidase domain has a characteristic α/β hydrolase fold [126, 149, 34, 145] containing a central eight-stranded β sheet (See Fig. 1.2). All of the eight strands except the second one are aligned parallel to each other. However, the β sheet itself is significantly twisted; it is parted into two wings by two helices on one side and six helices on the other side (Fig. 2.8) [116]. The propeller domain is based on a sevenfold repeat of four-stranded antiparallel β sheets. The β sheets are twisted and radially arranged around their central tunnel (Fig. 2.8 and Fig. 2.9) and facing each other. Their structural stability is sourced predominantly from hydrophobic interactions. The rest of the known propeller proteins differ in their structure from POP by having a closed loop (also known as the velcro) between their first and last blades [155]. The velcro will be explained with the G protein as an example. The G (guanine nucleotide-binding) proteins are a family of proteins that transmit chemical signals from outside the cells. In the β subunit of G proteins, the velcro is closed. The closure is caused by the hydrogen bonds in the main chain that connect one β sheet from the amino terminus and three antiparallel β sheets from the carboxyl terminus. In the six-, seven- and eight-bladed propellers the velcro is closed in a similar way [155, 14]. However, in smaller four-bladed proteins such as C-terminal domain in hemopexin and collagenase the Velcro is closed by forming a disulphide (SS-bond) bond between the first and last blades [51, 107]. In contrast, the POP does not stabilize the circular structure by this system: only hydrophobic interactions between the first and last blades keep the Velcro closed. The β-propeller is attached to the catalytic domain by the two connecting polypeptide main chains by forming hydrogen bonds and salt bridges but mainly with hydrophobic forces.
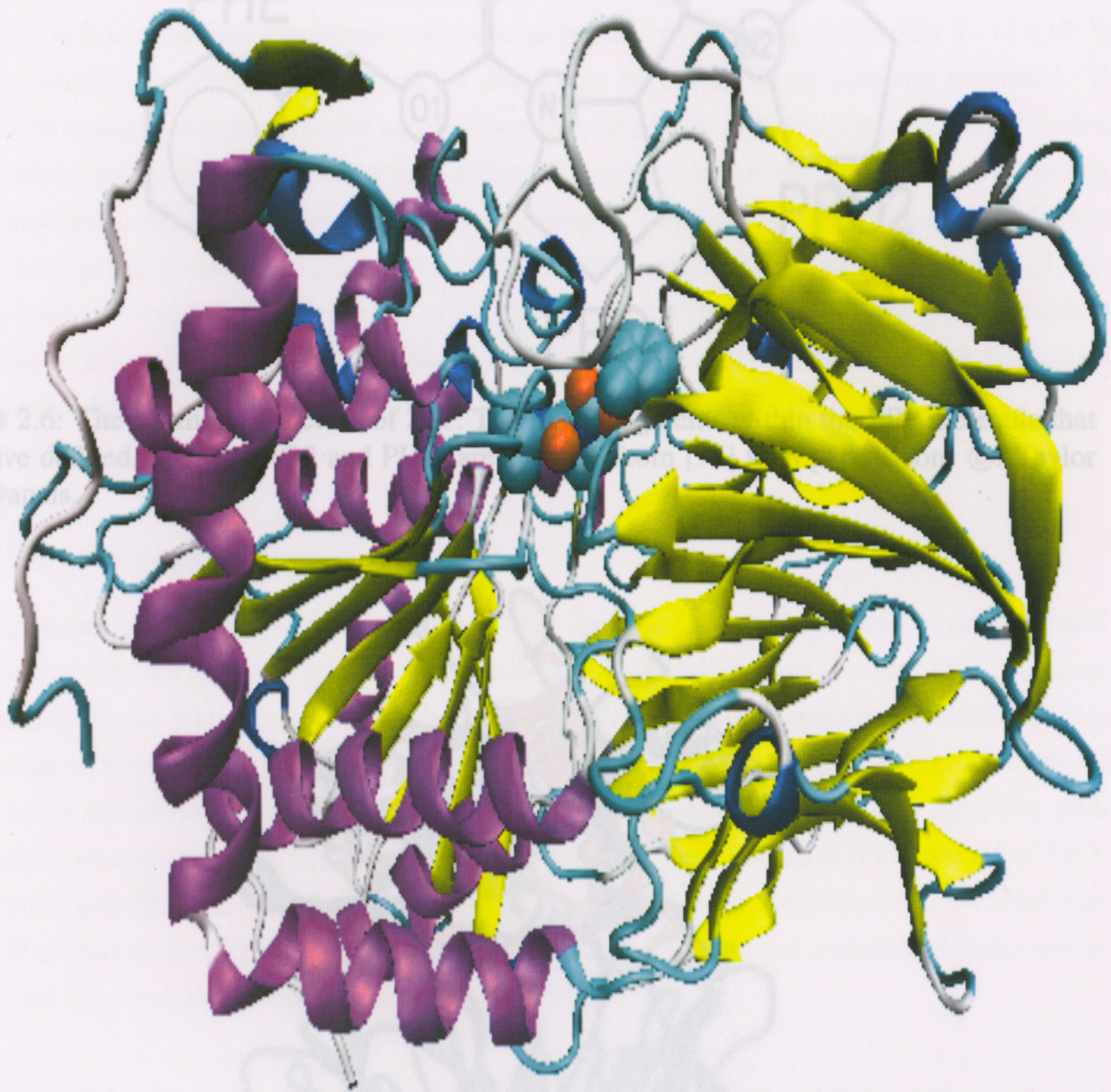
Figure 2.5: A snapshot of the structure of POP with ZPP front view, the protease catalytic domain is shown in purple and the seven-bladed β-propeller that acts as a gating filter mechanism is shown in yellow. The ZPP inhibitor inside the central cavity is bound to the protease catalytic domain through a reversible hemiacetal bond.

Figure 2.6: The chemical structure of ZPP. The three fragments within the ZPP molecule that we have defined, PRO1, PRO2 and PHE, are shown. From [95] with permission. ©: Taylor and Francis.
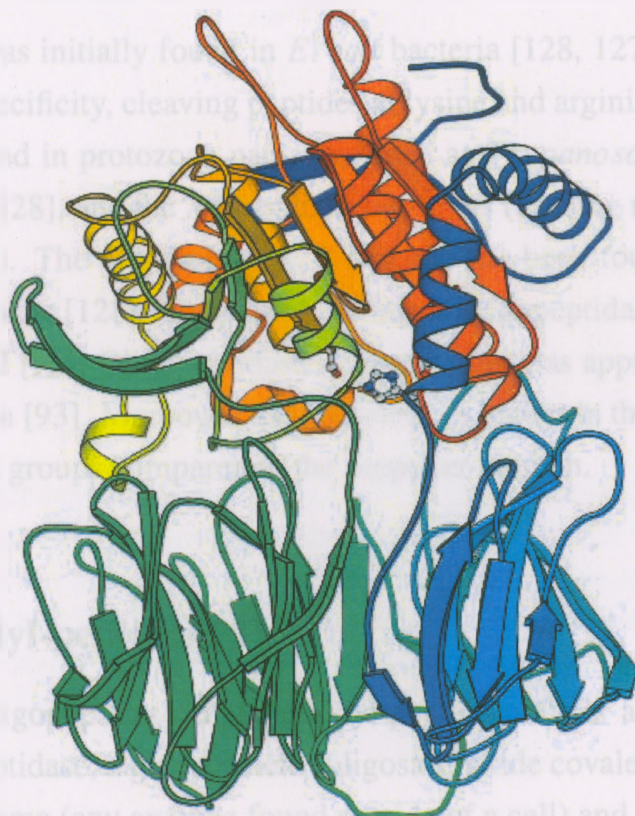
Figure 2.7: Structure of porcine POP. The ribbon diagram is color-ramped blue to red from the amino to carboxyl terminus. The catalytic residues are shown in ball and stick representation. From [132] with permission. ©: Springer

The active site composed of the catalytic triad (Ser554, Asp641, His680) is located in a large cavity at the interface of the catalytic and β propeller domains [55].

Proteases have different mechanisms in selecting substrates to prevent uncontrolled protein degradation. The entrance of the propeller in front of the active site is approximately $4 \times 10^{-12} m$ which is is very narrow compared to an average peptide which is approximately $6-12 \times 10^{-12} m$. The entrance is enlarged when blades 1 and 7 (the open blades) are partially separated. This system restricts the access to the active site preventing large peptides from accidental hydrolysis while still giving access to the rest of them. This mechanism was verified by synthesizing a disulphide bond between blades 1 and 7, resulting in inactivating the enzyme [56]. Moreover, the oscillating propeller blades on the enzyme act as a gating filter during catalysis [132]. It has been observed that the propeller of the hyperthermophilic enzyme (a prolyl oligopeptidase) opens to a large extent at high temperatures in order to hydrolyze proteins such as azocasein [65, 132, 64].

2.2.2 Oligopeptidase B

Oligopeptidase B was initially found in *E. coli* bacteria [128, 127] where it was associated to have trypsin-like specificity, cleaving peptides at lysine and arginine residues [128]. Oligopeptidase B is also found in protozoan parasites, such as *Trypanosoma cruzi* (related to Chagas disease in humans) [28], and the African trypanosomes (causing nagana in cattle and sleeping sickness in humans). The oligopeptidase B enzyme has been found to be a cytosolic protein causing osmotic shocks [128]. The catalytic triad of oligopeptidase B is composed of Ser532, Asp617 and His 652 [93]. Oligopeptidase B from *E. coli* has approximately 25% similarity to POP of porcine brain [93]. Moreover, the homology is higher in the peptidase domain specially around the catalytic groups compared to the propeller domain.

2.2.3 Dipeptidyl-peptidase IV

Unlike POP and oligopeptidase B, dipeptidyl-peptidase IV is a dimer (two proteins bound together), an exopeptidase, a glycoprotein (oligosaccharide covalently attached to protein side-chains), an ectoenzyme (any enzyme found outside of a cell) and bound to the cell membrane. Dipeptidyl-peptidase IV mainly cleaves dipeptides with proline residue. Also it cleaves the alanine from the amino terminus of oligopeptides [33, 80]. Its cellular localization and enzymatic properties are different from those of other dipeptidyl-peptidases. The specificity of dipeptidyl-peptidase IV is similar to that of dipeptidyl-peptidase II. In contrast, neither
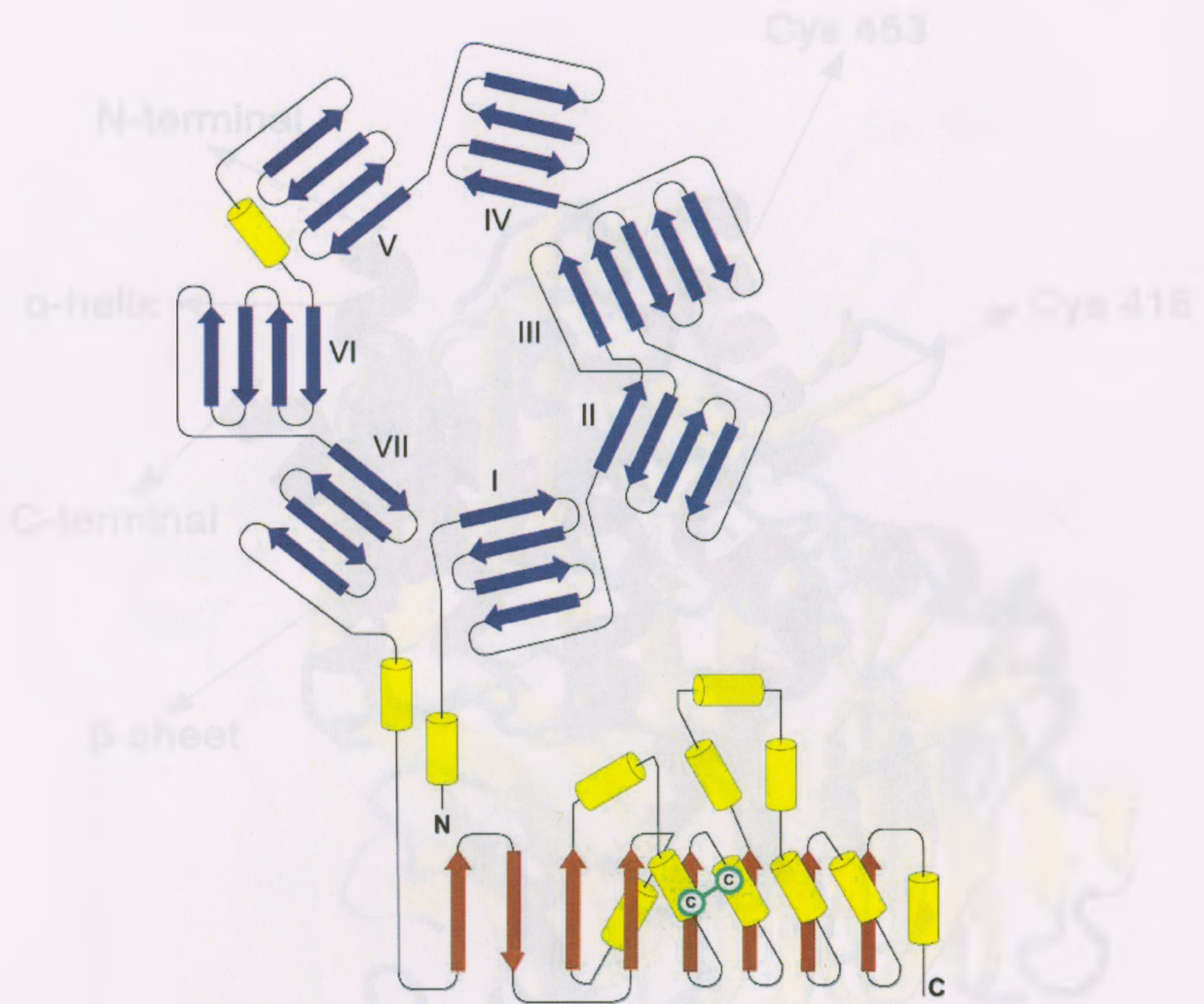
Figure 2.8: General structure of POPs family (e.g. acylaminoacyl-peptidase). The α helices are shown as yellow cylinders, β strands in the N-terminal domain are shown in blue, and β strands in the C-terminal domain are shown in red. The two cysteine residues in acylaminoacyl-peptidase (Cys416 and Cys453 of 1EV6.pdb) located in the C-terminal domain form a disulfide bridge between helices $\alpha6$ and $\alpha7$ and are shown as green circles. From [19] with permission. ©: Elsevier
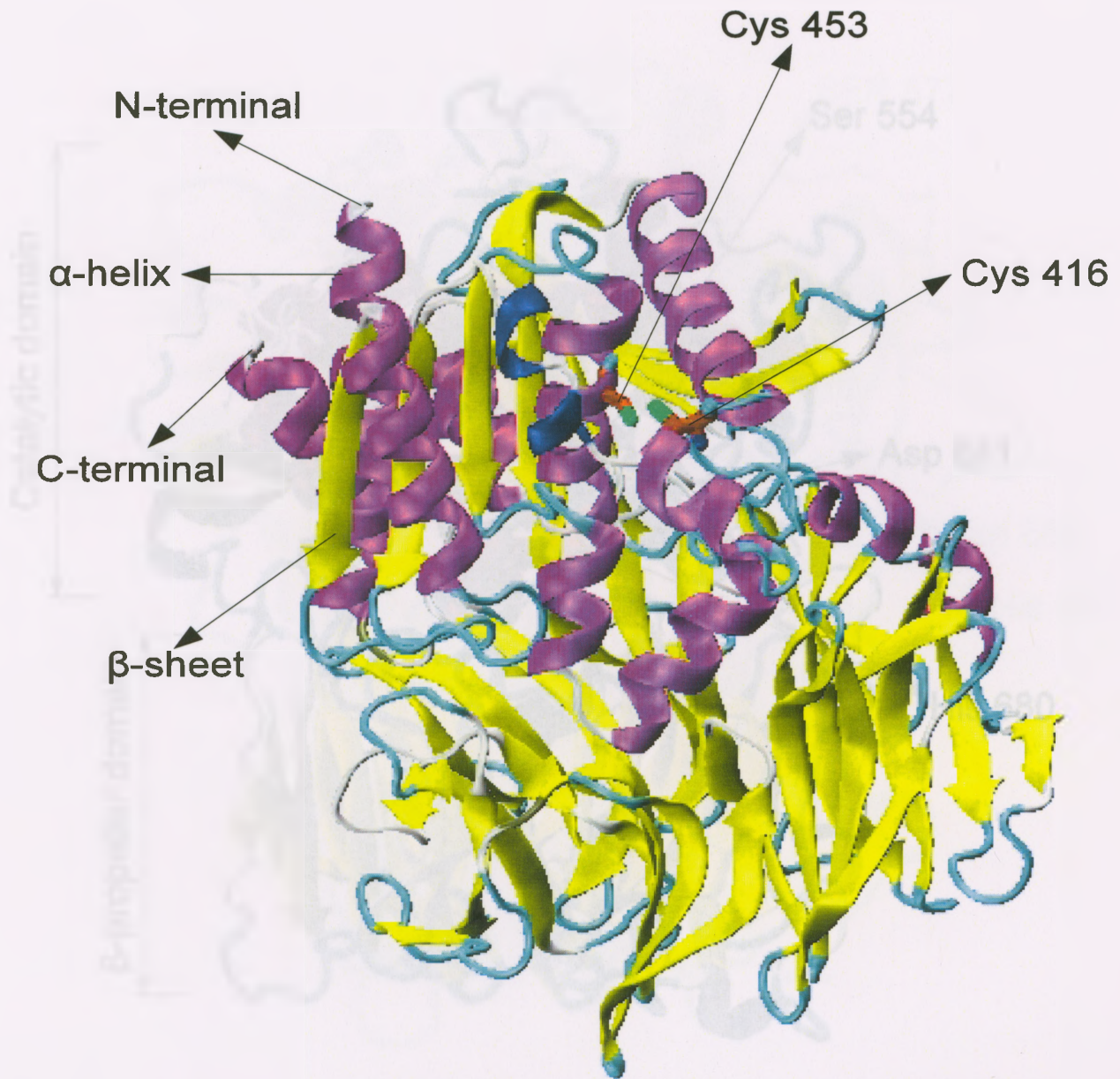
Figure 2.9: Two cysteine residues in Acylaminoacyl-peptidase (Cys416 and Cys453 of 1EV6.pdb [148]) α helix, β sheet, N-terminus and C-terminus are shown in this figure. In Cystein amino acids the green atoms are sulphur which make disulphid bond.
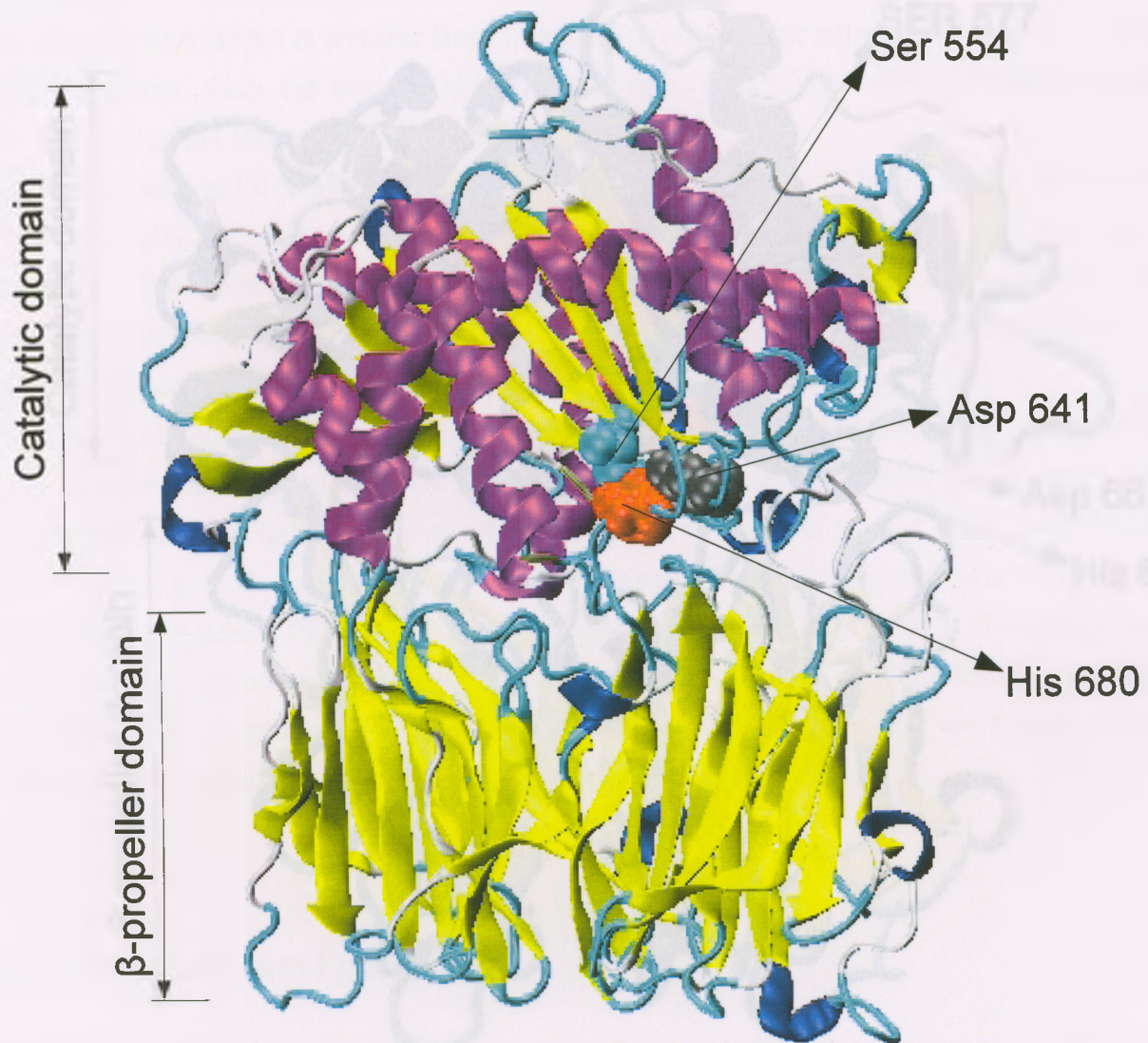
Figure 2.10: Side view of POP (1QFS PDB code [55]). The catalytic domain the β-propeller domain and the catalytic triad are shown
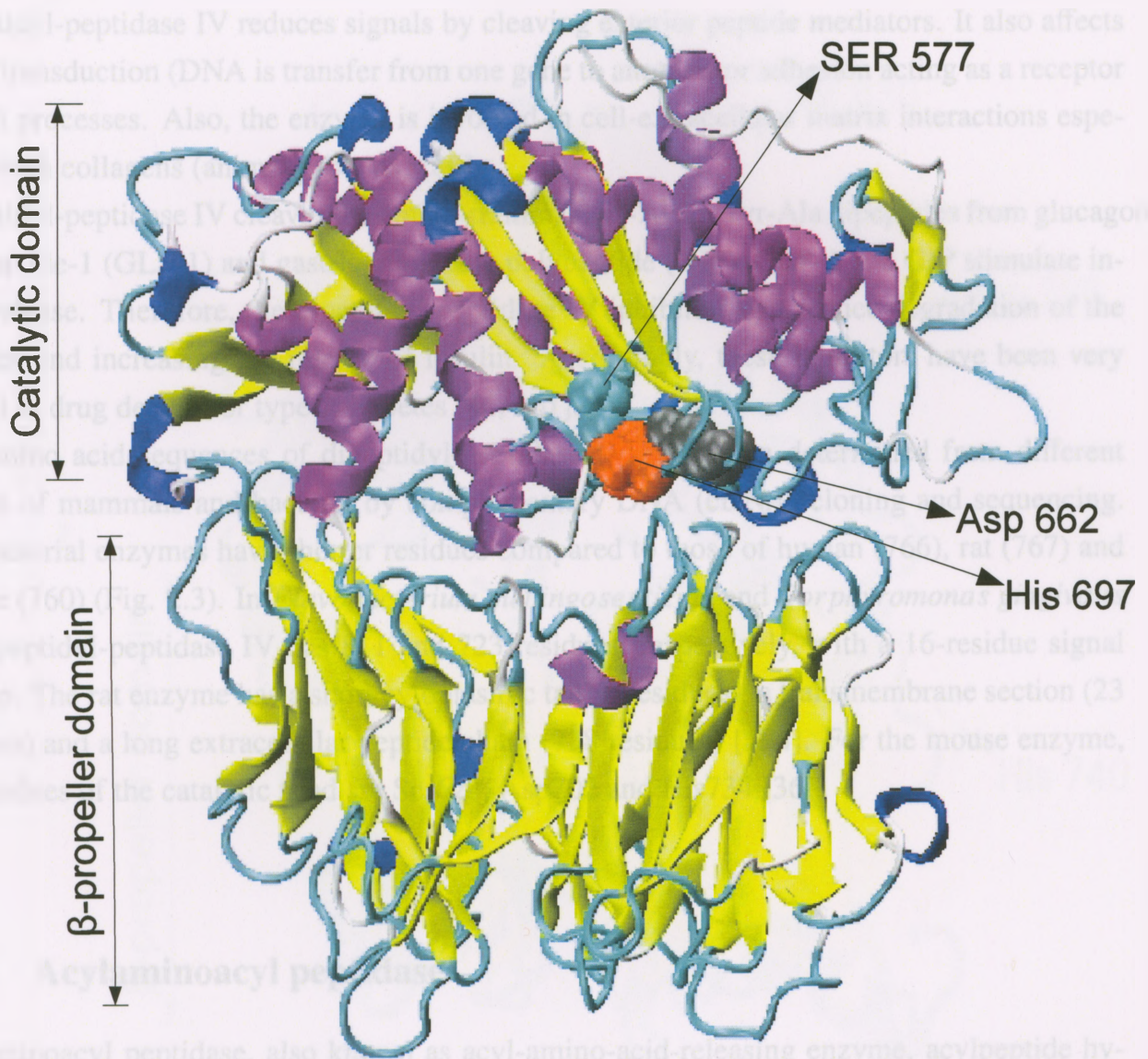
Figure 2.11: Side view of Oligopeptidase B(2XE4 PDB code [115]). The catalytic domain and β-propeller domain is shown also The catalytic triad are shown

dipeptidyl-peptidase I nor dipeptidyl-peptidase III cleave at proline residue [77].

Dipeptidyl-peptidase IV has many different functions in cells. In intestinal adrenal brush-border membranes, the dipeptidyl-peptidase IV affects peptide degradation and amino acid scavenging. In a study with rats lacking dipeptidyl-peptidase IV resulted in drastic weight loss to them [154].

Dipeptidyl-peptidase IV reduces signals by cleaving exterior peptide mediators. It also affects signal transduction (DNA is transfer from one gene to another) or adhesion acting as a receptor in both processes. Also, the enzyme is involved in cell-extracellular matrix interactions especially with collagens (animal proteins) [85].

Dipeptidyl-peptidase IV cleaves the amino-terminal His-Ala or Tyr-Ala dipeptides from glucagons like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP). GLP-1 and GIP stimulate insulin release. Therefore, the dipeptidyl-peptidase IV inhibitors can reduce degradation of the peptides and increasing the release of insulin. Accordingly, these inhibitors have been very helpful in drug design for type 2 diabetes [46, 161].

The amino acid sequences of dipeptidyl-peptidase IV have been determined from different species of mammals and bacteria by complementary DNA (cDNA) cloning and sequencing. The bacterial enzymes have shorter residues compared to those of human (766), rat (767) and murine (760) (Fig. 2.3). In *Flavobacterium meningosepticum* and *Porphyromonas gingivalis* the dipeptidyl-peptidase IV has 811 and 723 residues, respectively, with a 16-residue signal peptide. The rat enzyme has a short cytoplasmic tail (6 residues), a transmembrane section (23 residues) and a long extracellular peptide chain (738 residues) [124]. For the mouse enzyme, the residues of the catalytic triad are Ser624, Asp702 and His734 [36].

2.2.4 Acylaminoacyl peptidase

Acylaminoacyl peptidase, also known as acyl-amino-acid-releasing enzyme, acylpeptide hydrolase and acylaminoacyl peptide hydrolase, acts as a catalyst in the removal of an N-acylated amino acids from blocked peptides [88]. Acylaminoacyl peptidase cleaves peptides with different N-terminal acyl groups, such as acetyl, chloroacetyl, formyl and carbamyl [86]. Blocked peptides containing two or three amino acids are hydrolyzed faster than longer peptides [87], but the N-terminally blocked proteins are not a target of this enzyme. [52].

In contrast to POP and oligopeptidase B, which are monomers, and dipeptidyl peptidase IV, which is a dimer, acylaminoacyl peptidase has four identical subunits. It has been obtained from human erythrocytes [88], ovine liver [52], bovine liver [57], rabbit muscle [134], bovine lens [32] and porcine intestinal mucosa [135]. The sequences of human, porcine and rat acyl-
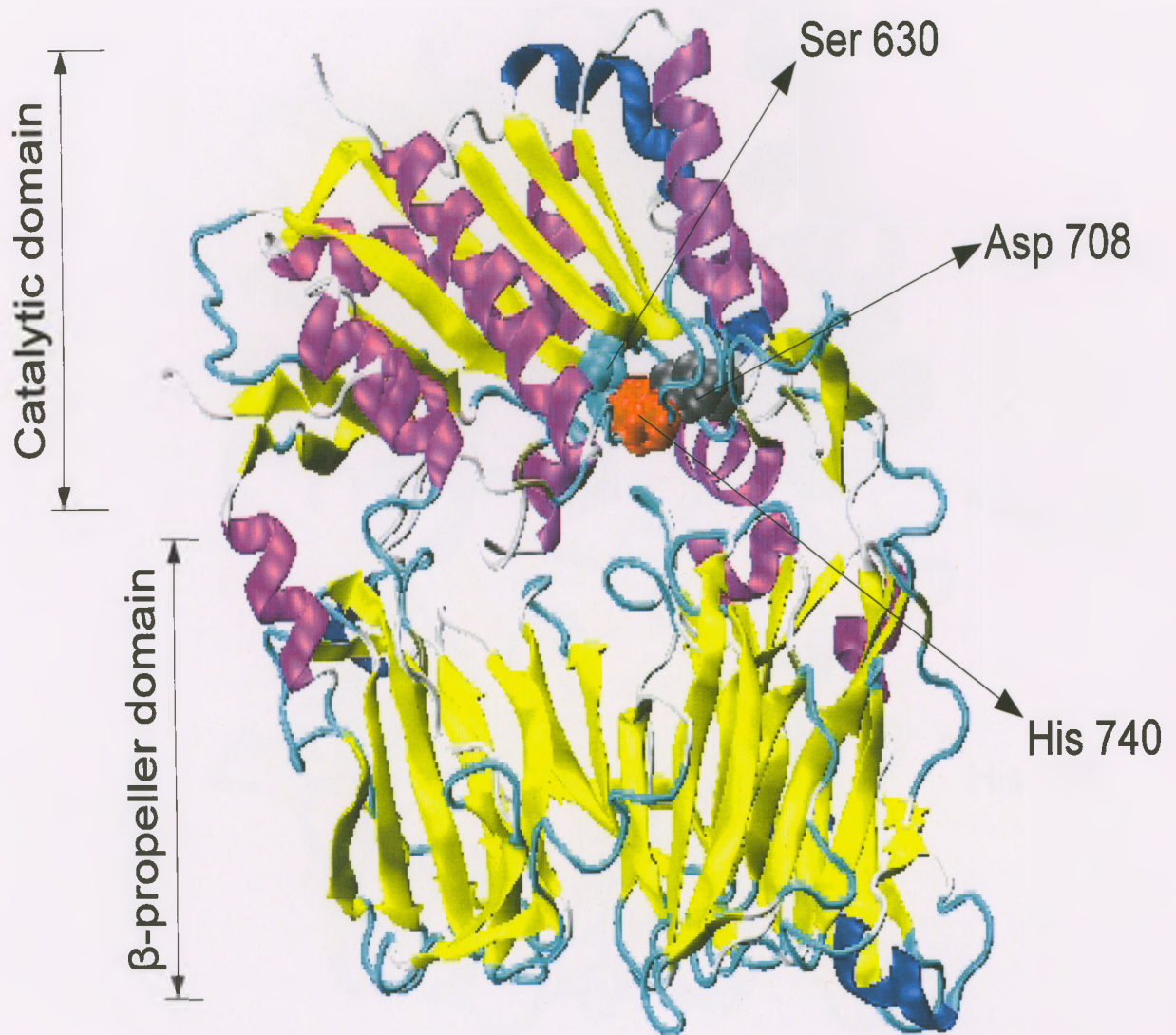
Figure 2.12: Side view of dipeptidyl-peptidase IV(1J2E PDB code [74]). The catalytic domain, the β-propeller domain and the catalytic triad are shown

aminoacyl peptidase have 732 residues and their sequences are over 90% identical to each other. The members of the catalytic triad are Ser587, His707 and Asp675 [132].

Acylaminoacyl peptidase has also been purified and cloned from the thermophilic archaeon *Pyrococcus horikoshii* [81]. It is found to be 100 residues shorter than the mesophilic acylaminoacyl peptidases, and is a dimer rather than a tetrameric enzyme [132, 81].
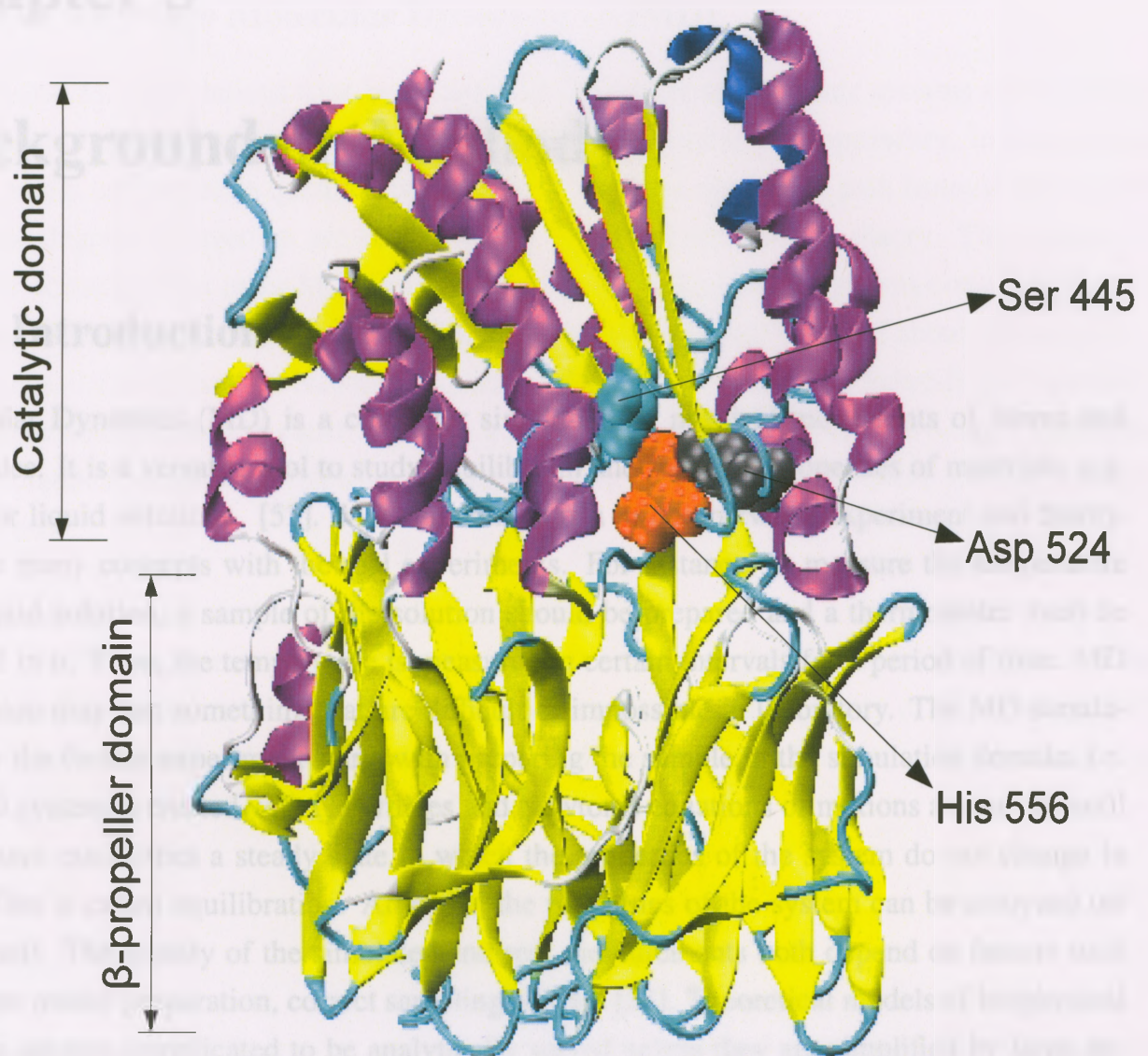
Figure 2.13: Side view of acylaminoacyl peptidases (1VE6 PDB code [20]). The catalytic domain, the β-propeller domain and the catalytic triad are shown

Chapter 3

Background on Methods

3.1 Introduction

Molecular Dynamics (MD) is a computer simulation of physical movements of atoms and molecules. It is a versatile tool to study equilibrium and transport properties of materials, e.g. alloys or liquid solutions. [53]. MD simulations is a bridge between experiment and theory. It share many concepts with the real experiments. For instance, to measure the temperature of a liquid solution, a sample of the solution should be prepared and a thermometer must be inserted in it. Then, the temperature is measured in certain intervals for a period of time. MD simulation may test something that are difficult or impossible in laboratory. The MD simulation for the former experiment starts with preparing the sample in the simulation domain, i.e. a model system is created with N particles and Newton's equations of motions are solved until the system establishes a steady state in which the properties of the system do not change in time. This is called equilibration. After that the properties of the system can be analyzed (or measured). The quality of the simulated and real measurements both depend on factors such as proper model preparation, correct sampling periods [32]. Theoretical models of biophysical systems are too complicated to be analytically solved unless they are simplified by large approximations.

The numerical simulations help to understand the behaviour of complicated theories and evaluate the parameters obtained from experiments. Numerical simulations resolve microscopic details of the system that are impossible or extremely hard to observe with experimental methods. Numerical simulations can also be used to validate and develop hypotheses made based on limited experiments. Computer simulations can be thought of as a tool to perform extremely controlled experiments, where every detail of the system can be adjusted. [63]

In the following sections, a summary of the different molecular dynamics approaches is presented.

3.2 Method

MD simulations can be divided into different categories depending on the method of calculating the forces within a given molecular system:

3.2.1 *Ab Initio* Molecular Dynamics (AIMD)

Ab initio MD simulations have been designed for studies of electronic systems since computational molecular dynamics system by using quantum mechanical approaches. In this approach the nuclei are treated as quantum particles with quantum mechanics path formula and the electronic degrees of freedom are treated based on electronic structure theory. This method has more accuracy than other MD simulations (Fig. 3.1). Simulation systems consist typically of about 10-1000 atoms and simulation times are limited to a maximum of about 0.1 ns [162].

Ab initio MD is based on Newton's equation of motion as well as the Schrödinger equation so this method reduces the equations from a fully quantum description (Schrödinger's equation) to a classical description (Newton's equation). One can divide the *ab initio* method into three sections:

1. *Born–Oppenheimer Molecular Dynamics*

   In Born–Oppenheimer MD, assume that the motions of electrons and nuclei can be separated from each other. The nuclei are propagated by integration of Newtons equation, means fixed nuclei position (nuclei is much heavier than electrons) at time and solve the equation stationary Schrödinger's equation:

   $$M_I \ddot{R} = -\nabla_I \min_{\Psi_0} \left\{ \left\langle \Psi_0 \left| \tilde{H}_{el} \right| \Psi_0 \right\rangle \right\}, \qquad (3.1)$$

   $$E_0 \Psi_0 = \tilde{H}_{el} \Psi_0 \qquad (3.2)$$

   where M_I and Ψ_0 are the mass of nucleus I and electronic ground state and E_0 is energy of the nuclei at ground state. $\tilde{H}_{el}$ is time-dependent via the nuclear coordinates $\{R_I(t)\}$. If we use KohnSham density functional theory to calculate E_0 when given KohnSham orbitals' (Φ_i are orthonormal) and position of 0th nuclear, an approximate electronic Hamiltonian ($\tilde{H}_{el}^{KS}$) must be calculated by iterative wave function optimization at each time step. The eigenvalue of $\tilde{H}_{el}^{KS}$ is calculated ($\tilde{E}_0$).

   Density functional theory (DFT) is a quantum mechanical modelling method used in evaluating the electronic structure of many-body systems (usually at the ground state). The Kohn-Sham DFT (KS DFT), is used to simplify the intractable many-body problem of interacting electrons in a static external potential [100, 129, 45]. The reduced prob-

lem only involves tractable non-interacting electrons moving in an effective potential the effective potential includes the external potential and the effects of Coulomb interactions between the electrons (including the exchange and correlation interactions).

2. Car–Parrinello Molecular Dynamics

The Car-Parrinello method [25] based on density functional theory is constructed to be based on the use of the electron density function. Because the Car-Parrinello method is MD combined with DFT in the adiabatic case, Car and Parrinello [25] introduced an extended Lagrangian which yields the forces achive on the electrons if differential to the possible orbitals interpreted as classical field, similar as for the classical nuclei. In Car–Parrinello MD simulations, the electronic density–functional theory is used to approximate the local density to compute the energies and densities of the electrons [138] or only of the valence electrons if the pseudo potential approximation is used. The Car–Parrinello method assumes that the system is in its electronic ground state and that electrons adiabatically follow the nuclear motions. Therefore motion of nuclei and electrons can be separated to maintain the system in the electronic ground state, the electronic energy must be minimized in each time step [26, 114].

3. Non–Born–Oppenheimer QM/MM molecular dynamics

A quantum mechanical (QM)/molecular mechanics (MM) framework implementation of an extended non-adiabatic AIMD approach can solve the problems related to size limitation of previously described methods in this chapter. In QM/MM framework, the electrons in the QM subsystem are described by a total wave-function ($\Psi^{QM/MM}$) which satisfies the time-dependent Schrödinger equation (TDSE) in environments with no photochemical reactions . The QM/MM coupling is defined through a Hamiltonian ($H^{QM/MM}$) which is a function of all the nuclear coordinates in the QM and the MM subsystems [102].

3.2.2   Classical Atomistic Molecular Dynamics

Classical MD simulations can be used to study the natural time evolution of classical many-body systems on a time scale of the order of 10^{-14}-10^{-8} second. The resolved state of the system for these time scales are suitable for the study of many structural and dynamical properties. However, the relevant fluctuations must diminish on time scales significantly shorter

Macroscale
Times: ~1 millisec and above
Lengths: over 1 micrometer
Example: anything visible to the naked eye
Methods: finite elements, phase fields

Mesoscale
Times: 10 ns to milliseconds
Lengths: 10 nm - 1 micrometer
Example: microphase separation of polymers
Methods: dissipative particle dynamics

Atomistic level
Times: ~nanoseconds
Lengths: ~nanometers
Example: molecular interactions
Methods: atomistic molecular dynamics

Subatomistic level
Times: femto to picoseconds
Lengths: over 1 micrometer
Example: electronic structure
Methods: ab initio methods

Simulated time scale increases

Simulation accuracy increases

Figure 3.1: Different time scales and length scales of typical computational methods and biological entities related to the various length scales. The dilemma between speed and accuracy is always present in simulations: it is always a trade-off between the two. In analogy to the various simulation methods at different time scales and length scales, it is not possible to use single experimental methods to cover all properties. Figure courtesy of Dr. M. Karttunen.

than the order of 10^{-8}s [53]. Classical MD simulations are able to reach time scales of the order of 100 ns.

Details of classical MD simulation in section 3.4.

3.2.3 Coarse Grained Classical Molecular Dynamics

Approximate coarse-grained models are used to study systems containing very many molecules for long of times. For example, colloidal suspensions involve dispersions of mesoscopic (10nm-1μm) solid particles (Fig. 3.1), where each of the particles consist of 10^6-10^9 of atoms [108, 53]. These make it nearly impossible to perform a classical MD simulation to observe several thousand colloids over an experimentally relevant time interval (milliseconds to seconds). Therefore, colloidal suspensions must be modelled using coarse-grained models.

The conventional all-atom methods for MD simulations need large amounts of computer memory and require long processing times. Moreover, the simulations for systems with long time scales (e.g. larger than10ns) would require extensive amount of processing due to the large number of timesteps the simulation should go through. Therefore, the reduced scale (coarse-grained) models should replace the all-atom methods [53], because in coarse grained method instead of using every atom of system (classical MD), it use a group of atoms [53]. Coarse-graning approaches are needed to study systems with length scales in 10^{-12} m- 10^{-9} m and time in 10ns- 0.01s (Fig.3.1) [159].

In some coarse graining methods, e.g. data partitioning [117], the derivation of interaction [17] potentials between CG particles is targeted at thermodynamic properties like energies and free energies. These methods are recommended for processes whose hydrophilicity or hydrophobicity arguments are more significant.

There are also structure−based coarse graining methods [110] in which the CG interactions are selected in a way that the model can generate specific structure properties that are commonly expressed by radial distribution functions. The distributions functions must be generated from all-atom molecular simulations [119, 109, 142]. Although the structure-based methods are suitable to re-insert atomistic coordinates, their performance in predicting thermodynamic properties and high-order structural correlations of the system is not clearly verified [141].

Another group of CG models uses a force matching method [48, 13] to predict CG force field to minimize the difference between the force in the atomistic system and the CG forces. This can be done by optimizing the CG interactions to create a many-body multi-dimensional potential of the mean force [122, 160, 141].

3.3   Integration

In order to solve the equations of motion which are nothing but differential equations, they must be properly discretized in time and space. The equations are integrated with respect to time. The accuracy and efficiency of the simulation is directly sensitive to the integration for Newton's equation of motion (eq. 3.3).

$$m_i \frac{d^2 \vec{x_i}}{dt^2} = \vec{F_i} = -\frac{\partial E\left(\left\{\vec{x_i}\right\}\right)}{\partial \vec{x_i}} \tag{3.3}$$

where E is the potential energy of the system that is obtained by the force–field (F_i force over ith particle), $\vec{x_i}$ distance from ith particle to others particle and dt time step.

For instance, the Verlet scheme in one dimension can be represented as:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + (\Delta_t)^2 \frac{\vec{F}(t)}{m} + O((\Delta_t)^4), \tag{3.4}$$

which indicates the scheme has accuracy of order of $(\Delta_t)^4$. The Verlet scheme for the equations of motion can be represented as:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \Delta t v(t) + \frac{1}{2}(\Delta t)^2 \frac{\vec{F}(t)}{m} \tag{3.5}$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{1}{2}\Delta t(\frac{\vec{F}(t + \Delta t)}{m} + \frac{\vec{F}(t)}{m}) \tag{3.6}$$

Leap-frog integration scheme is one of the most common schemes, which is a version of Verlet scheme. The leap-frog scheme for the equations of motion can be represented as:

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\vec{F}(t)}{m}\Delta t \tag{3.7}$$

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{3.8}$$

The leap–frog algorithm is time-reversible and preserves volume in phase space which are both important in Hamiltonian systems. Energy is conserved in short term and even the long term energy error is still small [75, 53, 121]. Since the MD simulations produce average properties of ensemble of particle, the exact trajectories and consequently the small drifts are less important. A disadvantage in leap–frog method is that the positions and velocities of

the particles are not calculated at the same time. However, the inconvenient related to this fact can be alleviated by averaging the velocities at plus and minus half a step and using the simultaneous velocities.

3.4 Force-Field

For an MD simulation the equations of motion must be solved using computational methods. As the initial setup the coordinates and momenta for all the particles in the system set up. Since in classical MD the quantum effects are not explicitly incorporated in the formulations, the time–dependent Newton's equations of motion are solved:

$$m_i \frac{d^2 \vec{x_i}}{dt^2} = \vec{F_i} = -\frac{\partial E\left(\{\vec{x_i}\}\right)}{\partial \vec{x_i}}, \qquad (3.9)$$

where E is the potential energy of the system that is obtained by the force–field ($\vec{x_i}$ is distance from ith particle to others particle).

The energy (E) is a function of the atomic positions (R) for all of the atoms in the system. The atomic positions are normally in Cartesian coordinates. Energy is decomposed into to two parts: bonded (internal) and non-bonded (external) energy. The bonded energy (E_{bonded}) corresponds to the bonds, angles and bond rotations in a molecule. These energy terms are typically described within a single molecule for atoms that are no further than 1, 2 or 3 covalent bonds away from each other, respectively. The non-bonded energy ($E_{non-bonded}$) represents the interactions between the atoms that do not share a bond or atoms that are separated by three or or more covalent bonds.

$$V(R) = E_{bonded} + E_{non-bonded} \qquad (3.10)$$

The E_{bonded} term (covalent term) is a sum of three terms:

$$E_{bonded} = E_{bond-stretch} + E_{angle-bond} + E_{rotate-angle-bond} \qquad (3.11)$$

bond strain: $E_{bond-stretch}$ (Fig. 3.2,a) is a harmonic potential representing the interaction between atomic pairs where atoms are separated by one covalent bond. It is the approximation to the energy of a bond as a function of displacement from the ideal bond length (r_{ij}^0). The force constant (k_{ij}^b) implies the strength of the bond. The ideal bond lengths (r_{ij}^0) and force constants (k_{ij}^b) are defined for each pair of bound atoms and depend on the type of atoms or constituents. The values for bonded interactions are obtained from experimental methods (X-ray crystallog-

raphy and spectroscopy) as well quantum mechanical calculations [15].

$$E_{bond} = \sum_{i,j} \frac{k_{ij}^b}{2} \left(r_{ij} - r_{ij}^0 \right)^2 \tag{3.12}$$

$$\tag{3.13}$$

The force constant is usually determined using experimental data, e.g. infrared stretching frequencies, or quantum mechanical calculations. Moreover, the bond length is determined from high resolution crystal structures or microwave spectroscopy data.

angle strain: $E_{angle-bond}$ (Fig. 3.2,b) is a harmonic potential related to deviations of the bond angles (θ) from the their ideal values (θ^0). The values of θ^0 and k_{ijk}^θ depend on the type of atoms constituting the angle. They actually represent the deviation from an ideal geometry. Since they are penalty functions, the sum of them should be close to zero in a perfectly optimized structure.

$$E_{ang} = \sum_{i,j,k} \frac{k_{ijk}^\theta}{2} \left(\theta_{ijk} - \theta_{ijk}^0 \right)^2 \tag{3.14}$$

$$\tag{3.15}$$

torsional potential: $E_{rotate-angle-bond}$ (Fig. 3.2,c) represents the torsion angle potential function. It models the presence of steric barriers between atoms separated by 3 covalent bonds. This term causes a rotation, described by a dihedral angle (φ_{ijkl}) and coefficient of symmetry ($n = 1, 2, 3$) (V_{nijkl}) around the middle bond. This potential is assumed to be periodic and is often expressed as a cosine function.

$$E_{rotate-angle-bond} = \sum_{i,j,k,l} \sum_n \frac{V_{nijkl}}{2} \left[1 + \cos \left(n\varphi_{ijkl} - \varphi_{ijkl}^0 \right) \right] \tag{3.16}$$

These energy terms are typically described within a single molecule for atoms that are no further than 1, 2 or 3 covalent bonds away from each other (dihedral angle), respectively.

The non-bonded interaction is sum of the LJ interaction energy (E_{LJ}) and the electrostatic interaction energy ($E_{electrostatic}$).

$$E_{non-bonded} = E_{LJ} + E_{electrostatic} \tag{3.17}$$

The LJ interactions are include repulsive and attractive forces between two atoms. The repul-
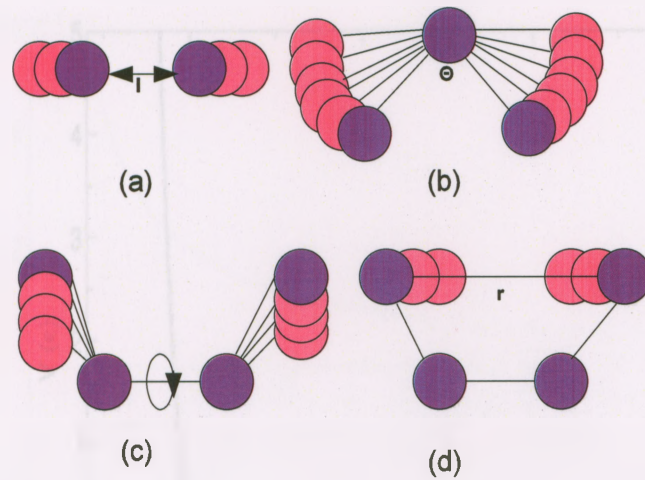
Figure 3.2: Ball-and-spring representations of (a) bond stretching, (b) angle bending, (c) torsion, and (d) nonbonded interaction.

sive force arises at short distances due to strong electron-electron interaction. The attractive force (dispersion force), originates from fluctuations in the charge distribution in the electron clouds. The fluctuation in the electron distribution on one atom or molecule creates a dipole that will induce another dipole on the other atom, which then leads to an attractive force. Both the repulsion and attraction effects are zero at long distances, but the repulsion force acts only at short distances while the attraction act at longer distances. Therefore, the atoms would stay at a distance which creates a minimum energy system. The value of energy at the minimum (ε) and the optimal separation of atoms (r_m) depend on the type of these atoms. Note that r_m is roughly equal to the sum of LJ radii of the atoms.

The LJ interaction is most often modelled using the Lennard-Jones (LJ) potential (Eq. 3.18). The LJ potential implements the interaction energy using the atom-type dependent constant C^6 and C^{12}. Values of C^6 and C^{12} are determined by experiments such as non-bonding distances in crystals and gas-phase scattering measurements. In the LJ interactions repulsion term is $\frac{C_{ij}^6}{r_{ij}^6}$ and dispersion term is $\frac{C_{ij}^{12}}{r_{ij}^{12}}$.

$$E_{Lennard-Jones} = \sum_{non-bonded\,pairs} \left(\frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6} \right) \tag{3.18}$$

The LJ potential may also be written as:

$$E_{Lennard-Jones}(\mathbf{r}_{ij}) = \sum_{ij} \left(\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} \right) - \left(\frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right). \tag{3.19}$$
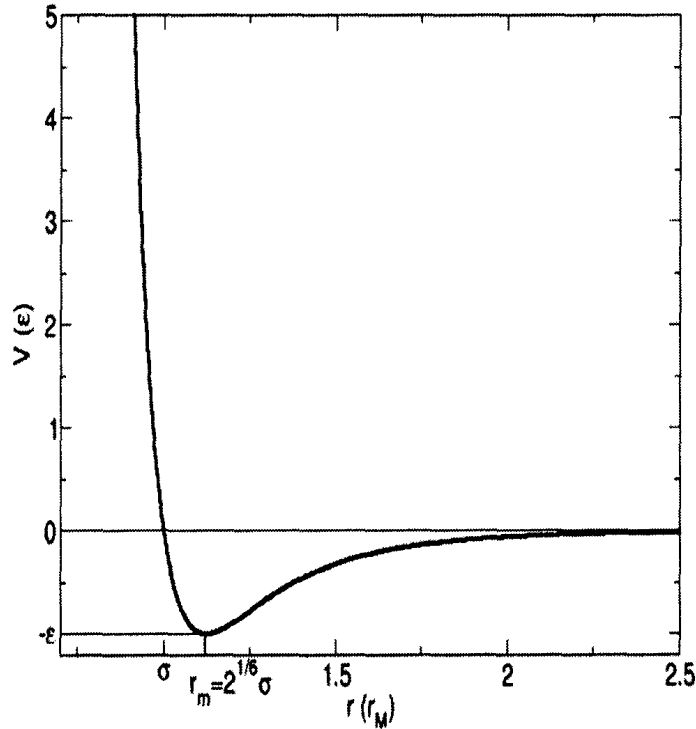
Figure 3.3: Lennard-Jones Interaction, ε is the depth of the potential and r_m is distance in that potential, σ is that distance which potential is zero. The figure is licensed under Creative Commons Attribution-Share Alike 2.5 Generic license. Source: Wikipedia [104].

The electrostatic interaction between a pair of atoms ($E_{electrostatic}$) is represented the electrostatic interactions between all the charged atom pairs by Coulomb potential:

$$U_{coul} = \sum_{i,j} \left(\frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \right) \tag{3.20}$$

The ε_0 is the electrical permittivity of space and r_{ij} is the distance between two atoms having charges q_i and q_j.

In my work we used cut-off radius for LJ interactions and Coulomb interactions. They can be modified by a shift function to replace the truncated forces by forces that are continuous and have continuous derivatives at the cut-off radius. The shift function produces a considerable modification of the Coulomb potential but for LJ potential is minor. Then coulomb potential was modified by using Particle-Particle Particle-Mesh (PPPM), Ewald, or PME [1]. In this thesis we used Particle-mesh Ewald (PME). So first would explain Ewald summation: In normal Ewald summation [32], a generic interaction potential (e.g. coulomb potential) ($\varphi(r)$) is separated into two terms: a short-ranged part $\varphi_{sr}(r)$ that sums quickly in real space and a long-ranged part $\varphi_{lr}(r)$ that sums quickly in Fourier space. The basic idea of particle mesh Ewald
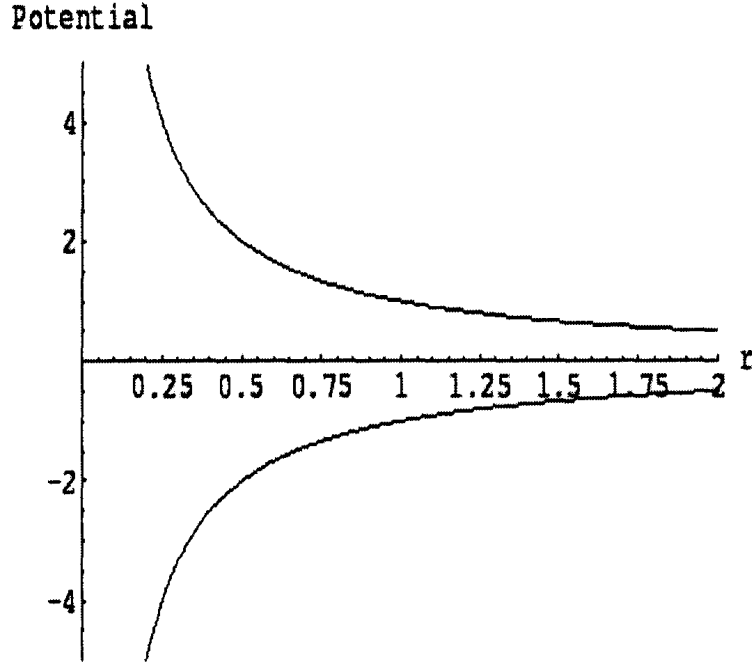
Potential

Figure 3.4: Coulomb potential; interaction between two particles of opposite charge(e.g. electrons) for lower part and upper part is for two same charge(e.g. electron and proton), r distance between two particles The figure has been released into the public domain. source: srikant.org [151].

summation is to replace the direct summation of interaction energies between point particles.

$$E_{total} = \sum_{i,j} \varphi(r_j - r_i) = E_{sr} + E_{lr} \qquad (3.21)$$

Where, E_{sr} is the sum of the short-ranged potential in real space (particle part):

$$E_{sr} = \sum_{i,j} \varphi_{sr}(r_j - r_i) \qquad (3.22)$$

and E_{lr} is a summation in Fourier space of the long-ranged part:

$$E_{lr} = \sum_{k} \tilde{\Phi}_{lr} \left| \rho(\tilde{k}) \right|^2 \qquad (3.23)$$

$\tilde{\Phi}_{lr}$ and $\rho(\tilde{k})$ are the Fourier transforms of the potential and the charge density.

Both of the above summations converge quickly (exponential convergence) [50, 41, 42, 103] in real and fourier spaces. The summations E_{sr} and E_{lr} converge with a high rate in real and fourier spaces so that they can be truncated without increasing the numerical error with lower the numerical effort. The Fourier transform $\rho(\tilde{k})$ of the charge density field is obtain by the Fast

Fourier transform (FFT) method by evaluating the density field on a discrete lattice in space. [103]

The final $E_{total-coulomb}$ would be as:

$$E_{total-coulomb} = \frac{1}{2V} \sum_{k \neq 0} \frac{4\pi}{k^2} |\rho(k)|^2 e^{\frac{-k^2}{4\alpha}} - (\frac{\alpha}{\pi})^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 + \frac{1}{2} \sum_{i \neq j}^{N} \frac{q_i q_j}{r_{ij}} \text{erfc}(\sqrt{\alpha} r_{ij}) \qquad (3.24)$$

that

$$E_{lr} = \frac{1}{2V} \sum_{k} \tilde{\Phi}(k) e^{ik.r} = \frac{1}{2V} \sum_{k \neq 0} \frac{4\pi}{k^2} |\rho(k)|^2 e^{\frac{-k^2}{4\alpha}}, \qquad (3.25)$$

$$E_{sr} = \frac{1}{2} \sum_{i \neq j}^{N} \frac{q_i q_j}{r_{ij}} \text{erfc}(\sqrt{\alpha} r_{ij}), \qquad (3.26)$$

$$\rho(k) = \sum_{k}^{N} q_i e^{ik.r}, \qquad (3.27)$$

also $E_{total-coulomb}$ includes a self interaction (a point charge interacts with charge cloud), so E_{self} will be:

$$E_{self} = -(\frac{\alpha}{\pi})^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 \qquad (3.28)$$

One of the limitations in these methods rises from using a fixed set of atom types to determine the parameters for the force field. Usually, the atom types are used to define an atom in a particular bonding situation. For instance, an aliphatic carbon atom will be different from a carbon atom found in a Histidine ring. However, in the simulations, instead of presenting each atom in the molecule as a unique one described by unique set of parameters, the atom types are grouped into certain number of sets to minimize the number of atom types and maximize simulation speed.

Tom Draden [49], for updating Ewald summation, proposed Particle-mesh Ewald [72]. PME uses a fixed cutoff in the direct sum also it uses B-spline interpolation of the reciprocal space structure factors onto a regular grid which obtain forces by analytic differentiation of the energies and it permit to use of fast Fourier transforms to calculate the reciprocal sum be efficient [49]. This method is arbitrarly accurate and scales as $N \log(N)$ with respect to system size instead of order of N^2 in Ewald summation and is faster than Ewald summation on medium to large systems [72].

3.5 Thermostat

There are two ensembles that require the temperature be constant: the canonical ensemble (NVT) and the isothermal-isobaric ensemble (NPT).

In the canonical ensemble (NVT), the number of particles (N), system volume (V) and temperature (T) are constant. In NVT ensemble, the energy of endothermic and exothermic processes is exchanged with a thermostat.

In classical MD simulations by using Newton's equations of motion the energy is constant and so the simulations are performed in the microcanonical ensemble (NVE ensemble), where the volume, the number of particles and the energy are controlled. But in experiments, the temperature is often keep constant instead of the energy.

Thermostats keep the temperature of the system constant so that NVT and NPT ensembles can be correctly created.

The following relation shows the instantaneous value of the temperature in terms of the kinetic energy and the particle momentum:

$$\sum_{i=1}^{N} \frac{|P_i^2|}{2m_i} = \frac{k_b T}{2}(3N - N_c) \tag{3.29}$$

where, P_i is momentum of atom the ith and N_c is the number of constraints and $3N - N_c$ represents the number degrees of freedom and K_B is the Boltzmann constant. The average temperature $\langle T \rangle$ is identical to the macroscopic temperature.

In the isothermal-isobaric ensemble (NPT), the number of particles (N), pressure (P) and temperature (T) are conserved. To obtain a NPT ensemble a barostat is needed in addition to a thermostat. An example of the NPT conditions is a flask open to ambient temperature and pressure.

A number of thermostats are available to add and remove energy from the boundaries of an MD system to approximate the canonical ensemble. The most common methods to control the temperature are velocity rescaling, the Nosé–Hoover thermostat, Nosé–Hoover chains, the Berendsen thermostat and Langevin dynamics [144].

The temperature control and evaluating the trajectories generated by molecular dynamics simulations are the most important issues that must be resolved [5, 53]. In classical molecular dynamics the microcanonical ensemble NVE is generated due to the conservation laws of Hamiltons equations. In NVE, the the number of particles, volume and the energy are kept constant. The most basic temperature control method is to rescale the velocities until the system is equilibrated at the target temperature. In addition, the energy conservation must be observed to ensure correct NVE ensemble sampling, to choose integration time-step and to

monitor numerical errors [24].

In contrast, Nosé thermostat controls the temperature without using a random number and also providing a conserved quantity. This has made Nosé thermostat very popular but it can exhibit non ergodic behavior.

Berendsen thermostat fixes the total kinetic energy by adding a first order equation for the kinetic energy to the Hamiltons equations [22]. This kinetic energy is originated from the difference between the instantaneous kinetic energy and its target value. Berendsens thermostat is stable and physically justifiable but it does not have a conserved quantity and is not associated to a well defined ensemble.

3.5.1 Velocity Scaling

Since most of the experiments are carried out under conditions that do not represent the microcanonical ensemble (NVE), in order to examine the behaviour of the system at a specific temperature, a NVT simulation should be performed with employing a thermostat. Moreover, when a thermostat is applied to a simulation, the steady energy drifts caused by the accumulation of numerical errors can be avoided. Velocity scaling can be used to keep the temperature of the system constant.

If the velocities are multiplied by a factor λ, the corresponding temperature change will be:

$$\Delta T = \frac{1}{2} \sum_{i=1} 2 \frac{m_i(\lambda v_i^2)}{N_{df} k_B} \tag{3.30}$$

$$\Delta T = (\lambda^2 - 1)T(t) \tag{3.31}$$

$$\lambda = \sqrt{\frac{T_0}{T(t)}} \tag{3.32}$$

where $T(t)$ is the temperature at time t.

Therefore, to control the temperature of the system one can multiply the velocities at each time step by the factor $\lambda = \sqrt{\frac{T_0}{T(t)}}$, where $T(t)$ is the current temperature calculated from the kinetic energy and T_0 is the desired temperature. However, in this approach the fluctuations in temperature present in the canonical ensemble are not accounted for.

Berendsen Temperature Coupling

The Berendsen thermostat [22] rescales the velocities of particles in the system to control the simulation temperature. In this method, a heat bath with the target temperature is weakly coupled to the system. Since, the thermostat reduces the fluctuations of the kinetic energy

of the system it does not produce trajectories, which are consistent with the microcanonical ensemble. With this thermostat, any deviations from the designed temperature exponentially decay with a temperature relaxation time (τ):

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{3.33}$$

where, τ is the coupling parameter (temperature relaxation time), controlling the coupling strength between the bath and the system. Therefore change in temperature between two time steps can be written as:

$$\Delta T = \frac{\delta t}{\tau}(T_0 - T(t)) \tag{3.34}$$

And the scaling factor for the velocities is:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left\{\frac{T_0}{T(t - \frac{\delta t}{2})} - 1\right\}. \tag{3.35}$$

The term $T(t - \frac{\delta t}{2})$ has emerged due to the usage of leap-frog algorithm integration in time. The value of τ should be chosen carefully, In practice, τ is obtained experimentally. With very large value τ ($\tau \rightarrow \infty$) the Berendsen thermostat is inactive and the simulation creates a microcanonical ensemble. With very small values of τ temperature fluctuations are suppressed reward unrealistic state. With τ being the same as the timestep δt, the Berendsen thermostat would just perform velocity scaling. Values of τ in the order of 0.1ps are typically used in MD simulations of condensed-phase systems.

It is worth to note that the thermostat does not generate a correct canonical ensemble for small systems but it can result in roughly correct approximations for large systems containing hundreds to thousands molecules. This scheme is efficient in adjusting the system temperature to the target temperature which makes it suitable for most simulations.

Commonly, the Berendsen scheme is used to take the system to an initial equilibrium state. Afterwards, the properties are calculated using Nosé–Hoover thermostat, that is capable of generating trajectories consistent with a canonical ensemble [79].

Parrinello and Bussi

The Parirrinello-Bussi thermostat is based on velocity scaling methods where velocities of all the particles are rescaled with a factor obtained from kinetic energy.

The method is also an extension of Berendsen thermostat with an added random force which

is constructed to implement the correct distribution for kinetic energy. With properly chosen relaxation time for the thermostat, the trajectories of the particles will remain approximately unchanged. While producing a correct canonical distribution, a quantity similar to energy in micro-canonical systems is available that can be monitored as a measure of simulation error. The Parirrinello-Bussi method starts with multiplying the velocities of all the particles by the same factor α, calculated from the total kinetic energy K and its target value $\bar{K}$:

$$\alpha = \sqrt{\frac{\bar{K}}{K}} \qquad (3.36)$$

$\bar{K}$ must be selected in such a way that the random changes in the kinetic energy leave the canonical distribution unchanged. The value for $\bar{K}$ can be selected from the previous value of K. First the system is updated for one time-step with Hamiltons equations using a time-reversible area-preserving integrator such as velocity Verlet [153]. Afterwards, the kinetic energy is calculated and marched one more time-step in time using an auxiliary continuous stochastic dynamics which can preserve canonical distribution:

$$\bar{P}(\bar{K})d\bar{K} \propto \bar{K}^{(\frac{N_f-1}{2})} e^{\beta \bar{K}} d\bar{K}, \qquad (3.37)$$

where N_f is the number of degrees of freedom and β is the inverse temperature. Finally, the the velocities are rescaled to enforce the new value of the kinetic energy [73].

The dynamics is described by a first-order differential equation in K. The auxiliary dynamics on K is one-dimensional and its associated Fokker-Planck equation [59] represents a zero-current solution:

$$dK = (D(K)\frac{\partial \log \bar{P}}{\partial K} + \frac{\partial D(K)}{\partial K})dt + \sqrt{2D(K)}dW, \qquad (3.38)$$

where $D(K)$ is an arbitrary positive definite function of K, dW is a Wiener noise. The distribution of Eq. 3.37 is inserted in the mention differential equation:

$$dK = \left(\frac{N_f D(K)}{2\bar{K}K}(\bar{K} - K) - \frac{D(K)}{K} + \frac{\partial D(K)}{\partial K}\right)dt + \sqrt{2D(K)}dW, \qquad (3.39)$$

This can be used to generate the correct canonical distribution. The above equation is independent on the choice of the function $D(K)$. However, $D(K)$ has effects on the speeds of equilibration. $D(K)$ is chosen to be

$$D(K) = \frac{2K\bar{K}}{N_f\tau}, \qquad (3.40)$$

where the arbitrary parameter τ has the dimension of a time and determines the time-scale of the thermostat such as in Berendsens formulation. This leads to a very transparent expression for the auxiliary dynamics

$$dK = (\bar{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{\bar{K}K}{N_f}}\frac{dW}{\sqrt{\tau}}. \tag{3.41}$$

Without the stochastic term this equation reduces to that of the standard thermostat of Berendsen. In the limit $\tau = 0$, the stochastic evolution is instantly thermalized and this algorithm reduces to a simple the stochastic velocity-rescaling method. On the other hand, for$\tau \to \infty$, the Hamiltonian dynamics is recovered. When a system is far from equilibrium, the algorithm leads to fast equilibration like the Berendsen thermostat [24].

Therefore cause of this method is a velocity rescales method, it is more ergodic than the Nosé–Hoover thermostat [24] and we used it for our work.

3.5.2 Nosé-Hoover Temperature Coupling

As mentioned before, the Berendsen thermostat is usually used to relax the system temperature to a target temperature in a few possible numerical iterations. To take the system to a correct canonical ensemble a method originally introduced by Nosé and developed by Hoover is implemented. Their method consists of a heat bath as an integral part of the system introducing $\tilde{s}$, associated with a mass $Q > 0$ as well as a velocity $\dot{\tilde{s}}$. Q determines the coupling between the reservoir and the system to control the temperature fluctuations, $\tilde{s}$ is a time-scaling parameter. The time scale in the extended system is increased by an amount of $\tilde{s}$.

$$d\tilde{t} = \tilde{s}dt \tag{3.42}$$

Since the atomic coordinates are identical in both systems:

$$\tilde{r} = r, \qquad \dot{\tilde{r}} = \tilde{s}^{-1}\dot{r}, \qquad \tilde{s} = s \quad and \quad \dot{\tilde{s}} = \tilde{s}^{-1}\dot{s} \tag{3.43}$$

The Lagrangian for the extended system is:

$$L = \sum_i \frac{m_i}{2}\tilde{s}^2\dot{\tilde{r}}_i^2 - U(\tilde{r}) + \frac{1}{2}Q\dot{\tilde{s}}^2 - gK_bT_0\ln\tilde{s} \tag{3.44}$$

The first two terms of the Lagrangian represent the kinetic energy and the potential energy of the real system, respectively. The additional terms are the kinetic energy of $\tilde{s}$ and the potential, chosen to ensure that the algorithm produces a canonical ensemble. The parameter g is chosen

$g = N_{df}$ in real-time sampling (Nosé–Hoover formalism) and $g = N_{df} + 1$ for virtual-time sampling (Nosé-formalism), N_{df} represents the system degree of freedom. This leads to the Nosé equations of motion.

$$\ddot{\tilde{r}}_i = \frac{\tilde{F}_i}{m_i \tilde{s}^2} - \frac{2\dot{\tilde{s}} \dot{\tilde{r}}_i}{\tilde{s}} \tag{3.45}$$

$$\ddot{\tilde{s}} = \frac{1}{Q\tilde{s}}\left(\sum_i m_i \tilde{s} \dot{\tilde{r}}_i^2 - g_b T_0\right) \tag{3.46}$$

These equations sample a microcanonical ensemble in the extended system $(\tilde{r}, \tilde{p}, \tilde{t})$.

Although, the energy of the simulation is kept constant, the energy in the real system is not constant. The fluctuations of $\tilde{s}$ along with the heat transfer occurring between the system and a heat bath keep the system temperature.

However, since the stretched time scale in the Nosé equations is a virtual variable, the dynamical properties of a system can not be extracted when this thermostat is used. Nosé and Hoover, reformulated the the Nosé equations in terms of real system variables:

$$s = \tilde{s}, \qquad \dot{s} = \tilde{s}\dot{\tilde{s}}, \qquad \ddot{s} = \tilde{s}^2 \ddot{\tilde{s}} + \tilde{s}\dot{\tilde{s}}^2 \tag{3.47}$$

$$r = \tilde{r}, \qquad \dot{r} = \tilde{s}\dot{\tilde{r}}, \qquad \ddot{r} = \tilde{s}^2 \ddot{\tilde{r}} + \tilde{s}\dot{\tilde{r}}^2 \tag{3.48}$$

and with substituting

$$\gamma = \frac{\dot{s}}{s} \tag{3.49}$$

the Lagrangian equations of motion can be written as

$$\ddot{r}_i = \frac{F_i}{m_i} - \gamma r_i \tag{3.50}$$

$$\dot{\gamma} = \frac{-k_b N_{df}}{Q} T(t)\left(\frac{g}{N_{df}}\frac{T_0}{T(t)} - 1\right) \tag{3.51}$$

In both of the algorithms, the virtual mass Q and extended system energy E_e should be carefully handled.

Very large values for Q (loose coupling) may result in inadequate temperature control since the thermostat will simulate a microcanonical ensemble when $Q \to \infty$.

In theory, any finite positive mass should lead to the generation of a canonical ensemble, but if Q becomes too large, too many simulation iterations will be required to obtain the canonical distribution.

On the other hand, very small values for Q (tight coupling) may cause high-frequency tem-

perature fluctuations. $\tilde{s}$ may oscillate at a very high frequency which takes it away from the resonance frequency of the real system leading in decoupling it from the physical degrees of freedom (slow exchange of kinetic energy).

As a more realistic choice for the coupling strength, the Nosé equations of motion can be expressed as:

$$\dot{\gamma} = \frac{1}{\tau_{NH}} \left(\frac{g}{N_{df}} \frac{T_0}{T(t)} - 1 \right) \tag{3.52}$$

with the effective relaxation time being:

$$\tau_{NH}^2 = \frac{Q}{N_{df} k_b T_0} \tag{3.53}$$

The relaxation time can be estimated when calculating the frequency of the oscillations for small deviations $\delta\tilde{s}$ from the average $\langle\tilde{s}\rangle$ [123, 79, 76].

Chapter 4

Structural Properties

4.1 Introduction

In this research GROMACS [150] (GROningen MAchine for Chemical Simulations) version 4.5.3 [1, 71] was used. GROMACS was developed at the University of Groningen, The Netherlands, in the early 1990s. This is well suited for parallelization on processor clusters. It is a very fast program for molecular dynamics simulation. It does not have a force field of its own, but is compatible with GROMOS [30], OPLS-AA [90], AMBER [27], and ENCAD [105] force fields [150]. The protein structure studied in the present work is Prolyl oligopeptidase from porcine muscle with covalently bonded inhibitor Z-Pro-Prolinal (ZPP) (PDB code 1QFS [54]). Here we first simulate POP protein alone, then put ZPP inhibitor in the cavity tunnel of POP and simulate POP with non-bonded ZPP.

All atom types of ZPP and POP are accessible in the OPLS [89] force field. OPLS-AA/L all-atom force field (Optimized Parameters for Liquid Simulations) are functions potential developed for proteins.

For water, the TIP3P (transferable intermolecular potential 3P) [91] model was used. Partial charges on ZPP were taken from previous study by K. Kaszuba *et al.* [95]. The Partial charges at ZPP were calculated by fitting electrostatic potential of the molecule using the restrained electrostatic potential (RESP) method [21]. The electronic structure and the electrostatic potential were calculated with the $6 - 31G^*$ basis set [143], compatible with the OPLS force field. A unit cell containing the protein and ZPP was defined with the dimension of $10 \times 10 \times 10$ nm^3 which was filled with water. This create a solvated system that contains a charged protein. In order to neutralize the system and create buffer saline conditions (140 mM salt concentration), potassium and chloride ions were added. Potassium was used instead of sodium to achieve a more realistic simulation of the conditions inside a living cell where potassium is the dominant anion [95]. Although the difference between K^+ and Na^+ may seem negligible at this level of
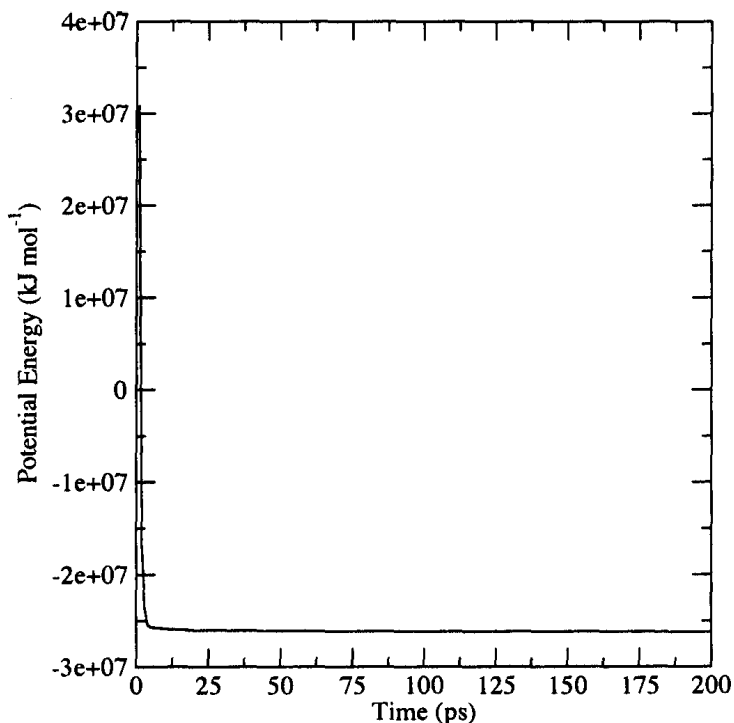
Figure 4.1: Potential energy of the system is minimized to sure each particle are in the right position

description, it has been shown that even the parameterization of the atomic-level ion force field is a topic of concern [11, 130].

Before starting the dynamics, the system must be checked to ensure there are no steric clashes or inappropriate geometry. The structure was relaxed through a process called energy minimization (EM). Prior to MD simulations, the energy of the structure was minimized using the steepest-descent algorithm (200 steps) in order to remove close contact between atoms which occurs after removing the covalent bonds. When the system is at an energy minimum (Fig. 4.1), the MD simulation can be started.

To begin MD, the solvent and ions must be equilibrated around the protein. Running unrestrained dynamics at this point may cause the system to collapse. The reason is that the solvent is mostly optimized within itself, and not necessarily with the solute. The solvent temperature must be taken to the desired level set in the simulation. Also it should establish the proper orientation about the solute (the protein). When the system is taken to the correct temperature, pressure should be applied to the system until it reaches the proper density. In equilibrating the protein-ligand complex (POP with bonded ZPP), ZPP is added to restraints of POP and temperature is added coupling groups (is controlled to be constant).

Equilibration is often conducted in two phases. The first phase is conducted under an NVT ensemble (constant number of particles, volume, and temperature). This ensemble is also re-
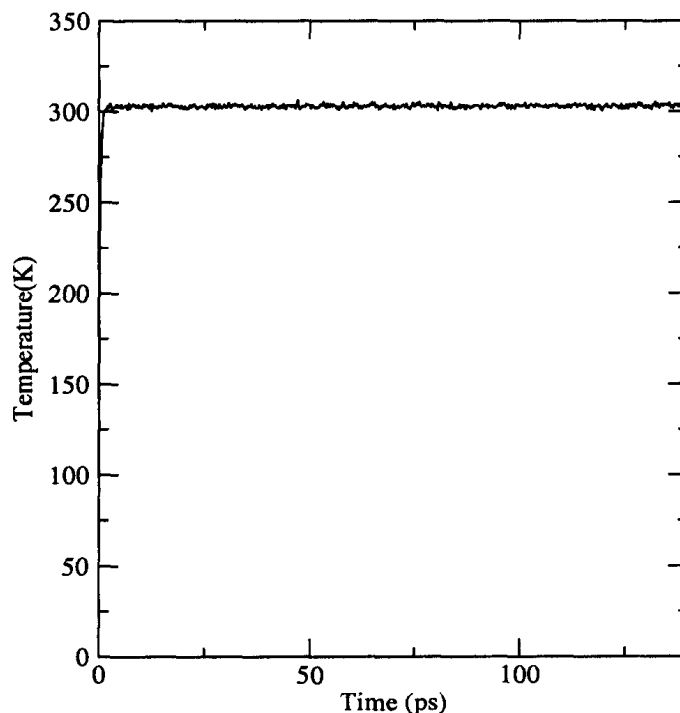
Figure 4.2: Temperature of the system in NVT equilibration: Temperature of the system quickly reaches the target value (303 K).

ferred to as isothermal–isochoric or canonical. In NVT, the temperature of the system should reach a plateau at the desired value. If the temperature has not yet stabilized, additional time is required. We will conduct a 140ps NVT equilibration to achieve this state.

Figure 4.2 shows that the temperature of the system quickly reaches the target value (303 K), and remains stable over the remainder of the equilibration.

The previous step, NVT equilibration, stabilized the temperature of the system. Prior to data collection, we must also stabilize the pressure (and thus also the density) of the system. Equilibration of pressure is conducted under the NPT ensemble, where the number of particles, pressure, and temperature are all constant. The ensemble is also called the isothermal-isobaric ensemble, which closely resembles experimental conditions.

After completion of the two equilibration phases, the system is well-equilibrated at the desired temperature and pressure. At this point the position restraints can be released production MD can be started for data collection.

4.2 Simulation details

The initial structure for POP with non-bonded ZPP was taken from K. Kaszuba *et al.* [95] where was performed MD study of the ZPP inhibitor non-bound in the binding cavity with the

POP was studied for 100ns. The present work focuses on two sets of simulation for the protein POP. The first set covers the protein itself and the second set is the protein (POP) with ZPP inhibitor.

At the first stage, the ZPP inhibitor was removed from the original POP with ZPP simulations provided by K. Kaszuba *et al.* [95]. Afterwards, the system was simulated for 50 ns more. Then the ZPP inhibitor was added to POP and the system was equilibrated, then the simulation for POP with non-bonded ZPP was continued for 50 ns. However, the charge group for ZPP was modified from the original version in K. Kaszuba *et al.* [95] for compatibility issues with the newest GROMACS version (4.5.3).

For the MD simulations, the POP molecule was immersed in a box of water with the dimensions of $10 \times 10 \times 10$ nm. The solvated simulation box contained a total of 99099 atoms (11219 atoms are protein, 80 ions are Cl ions, 85 ions are K and the rest of atoms are water atoms) for POP without ZPP and 98599 atoms for POP with ZPP inhibitor (11219 atoms are protein, 46 atoms are ZPP inhibitor, 80 ions are Cl, 85 are K ions and the rest of atoms are water).

The following parameters were used for the simulation. The algorithm used for integrating Newton's equations of motion was the leap-frog algorithm. The first stage of the simulation started at 0 time with 2 fs timesteps. The integrations were taken for each timestep and continued until time 50 ns for all simulations. Periodic boundary conditions with the usual minimum image convention were used in all three directions. Cut-off distance for the short-range neighbor list is 1.0 nm. Fast Particle-Mesh Ewald [35] electrostatics were employed for electrostatics. Distance for the Coulomb force cut-off was set to 1.0 nm.

Simulations were performed at constant temperature of 303 K. The V-rescale thermostat (See Section 3.5.1: Parrinello and Bussi) was used to keep the system temperature constant.

The LINCS (LINear Constraint Solver) algorithm [70] was used to reset bonds to their correct lengths after an unconstrained update.

The POP, ZPP, water, potassium, sodium and chlorine were coupled separately to the temperature bath. Time constant for temperature coupling was set to 0.1 K.

Simulations were performed with Parrinello-Rahman barostat which includes Extended-ensemble pressure coupling where the box vectors are subject to an equation of motion. The equation of motion for the atoms is coupled to this barostat. No instantaneous scaling takes place. The pressure coupling time constants was set to 0.5 kJ mol^{-1} nm^{-3}.

4.3 Measurements

We will study changes in the structure over the whole simulation time. There are two indicators for large structural changes in protein: the radius of gyration and the root mean square deviation

(RMSD).

4.3.1 Radius of Gyration

To analyze structural changes of a protein and have a rough measure for the compactness of a structure, one can calculate the radius of gyration of the structure (R_g). If a protein is stably folded, it will likely maintain a relatively steady value of R_g. If a protein unfolds, its R_g will change over time. R_g is defined as:

$$R_g^2 = \frac{1}{N} \sum_{i=1}^{N} (\vec{r}_i - \vec{r}_{mean})^2.$$ (4.1)

where,

$$\vec{r}_{mean} = \frac{1}{N} \sum_{i=1}^{N} \vec{r}_i$$ (4.2)

is the centre of mass of the molecular structure (it can be protein, backbone, C-alpha).

4.3.2 Root Mean Square Deviation

The root mean square deviation (RMSD) of certain atoms in a molecule with respect to a reference structure can be calculated by least-square fitting the structure to the reference structure (i.e. structure at $t = 0$). RMSD at time t, is calculated by

$$RMSD(t) = \sqrt{\frac{\sum_{k=1,N}(r_k - r_k(t = 0))^2}{N}}.$$ (4.3)

where, $r_k(t)$ is the position of the *kth* particle at time t. Least-square fitting does not have to use the same atoms as those used in the calculation of the RMSD. For example a protein is usually fitted on the backbone atoms (N, C-alpha, C), but the RMSD can be computed for the backbone or the whole protein. In this calculation, we use the RMSD of heavy-atoms to compare the spatial deviation between structures in time with the structure at reference time (t = 0 ps.) Typically, the RMSD of protein should not change more than 3^{-12} within a nanosecond of MD simulation time [1].

Our analysis also includes the fluctuations in the RMSD (RMSFs). When a dynamical system fluctuates about some well-defined average position the RMSD from the average over time is referred to as the RMSF or root-mean-square fluctuation.

The trajectories of the simulations were also visualized using the VMD (Visual Molecular Dynamics) package [78]. Also, the normal modes visualization are obtained with the help of Dynatraj online software [18].

4.4 Normal Modes

Normal mode analysis (NMA) is a simulation technique to analyze large-scale motions of biological molecules with complex shape changing motions [61, 106, 23]. This method was usually used in the experimental techniques of infrared and Raman spectroscopy. NMA is recently being used to predict the functional motions of large molecules such as proteins. The motions in the protein that perform a task for the molecule (e.g. binding protein to other molecules) are functional motions.

NMA is essentially a form of harmonic analysis. The purest form the NMA provides accurate results by using exactly the same force fields as used in molecular dynamics simulations. However, at the physiological temperatures, one of the assumptions in the method is not always correct. The conformational energy surface at an energy minimum can not always be approximated by a parabola over the range of thermal fluctuations at living body temperatures. It has been shown that the harmonic approximation breaks down in a way that the state point reaches multiple minimum states with different energies (instead of performing a harmonic motion in a single energy minimum) [12, 47]. At the functioning temperatures, these limitations in the assumptions must be carefully considered.

Performing an NMA involves three major steps. First, if the analysis is done in Cartesian coordinates, the conformational potential energy must be minimized as a function of the atomic Cartesian coordinates. Next, the second derivatives of the potential energy with respect to the mass-weighted atomic coordinates must be calculated in form of a matrix, called the Hessian matrix. Finally, the Hessian matrix must be diagonalized to obtain its eigenvalues and eigenvectors (the normal modes). The first and last steps involve extensive numerical calculations which can be very extensive for large molecules. The diagonalization process demands CPU time and large amount of memory because it involves diagonalization of a $3N \times 3N$ matrix, where N is the number of atoms in the molecule.

Theory

NMA is usually performed in vacuum where the potential energy of N-atom molecule is a complex function of its $3N$ coordinates. The potential function is usually written with respect to the Cartesian coordinates [23] composed of bonded and nonbonded energy terms. At a

minimum, the potential energy function U can be expanded in a Taylor series in terms of the mass-weighted coordinates, $\sqrt{m_i}\Delta x_i$, where Δx_i is the displacement of the ith coordinate from the energy minimum and m_i is mass of the corresponding atom. Knowing that the linear term is zero at an energy minimum, and writing the expansion up to the quadratic level, U will be:

$$U = \frac{1}{2} \sum_{i,j=1}^{3N} \frac{\partial^2 U}{\partial q_i \partial q_j}\bigg|_0 q_i q_j. \tag{4.4}$$

The above relation shows that the energy surface can be approximated by a parabola character-ized by the second derivatives evaluated at the energy at the minimum. The basic assumption of NMA of biomolecules at physiological temperatures is that the fluctuations occur within this parabolic energy surface. However, at these temperatures the state point moves on a complex energy surface with multiple minima, and reaches different energy levels [12]. The second derivatives in Eq. 4.4 form the Hessian matrix. By diagonalizing the Hessian, its eigenvectors and eigenvalues can be obtained:

$$\mathbf{F}\mathbf{w}_j = \omega_j^2 \mathbf{w}_j. \tag{4.5}$$

where $\mathbf{w}_j$ is the jth eigenvector and ω_j^2 is the jth eigenvalue. The above equation leads to 3N eigenvector equations, where each eigenvector specifies a normal mode coordinate using:

$$Q_j = \sum_{i=1}^{3N} w_{ij} q_i. \tag{4.6}$$

Note that the sum is over the elements of $\mathbf{w}_j$ and also $|\mathbf{w}_j| = 1$. These normal mode coordinates oscillate harmonically and independently of each other each with the angular frequency, ω_j. Therefore, if A_j is the amplitude and ε_j is the phase, the harmonic motion can be written as:

$$Q_j = A_j \cos(\omega_j t + \varepsilon_j). \tag{4.7}$$

These normal mode coordinates are linear combinations of the atom based Cartesian coordi-nates, as shown in Eq. 4.6. Considering a single normal mode, j, the displacements are:

$$\Delta x_{ij} = \frac{w_{ij}}{\sqrt{m_i}} A_j \cos(\omega_j t + \varepsilon_j) \tag{4.8}$$

Therefore, in the jth mode, the relative displacements of the Cartesian coordinates are specified by the elements of $\mathbf{w}_j$. This implies each normal mode specifies a pattern of atomic displace-ment. For example, in a multidomain protein, this pattern of displacement can represent the

relative movement of two domains in the molecule [67].

If the frequency of the oscillations is lower, the fluctuations of the corresponding normal mode coordinate are larger [60]. Usually the lowest frequency modes are compared with functional modes obtained from experimental methods (e.g., a pair of x-ray structures, one bound to a functional ligand and the other unbound). The overlap with the jth mode can be defined as [113]:

$$O_j = \frac{\sum_{i=1}^{3N} (\Delta x_{ij} \Delta x_i^{exp})}{\sqrt{\sum_{i=1}^{3N} (\Delta x_{ij})^2} \sqrt{\sum_{i=1}^{3N} (\Delta x_i^{exp})^2}} \tag{4.9}$$

Computing the Normal Modes

As mentioned before, the first step to an NMA is energy minimization. In some cases, the actual minimum energy cannot be found because of overstepping. This can present a problem for NMA, where very precise location of the minimum is required. The Hessian matrix must be calculated and then diagonalized in order to determine the eigenvalues and eigenvectors. Because of the large size of this matrix ($3N \times 3N$ matrix, N being the number of atoms in the molecule), this stage often presents memory problems for large molecules. The NMA results are usually compared with the experiments. The overlap with a functional mode derived from experiments (such as two x-ray structures) are calculated by Equation 4.9. The experimental displacement (Δx_i^{exp}) needs to be calculated from the experimental structures that are oriented in the same way as the minimized structure used for the NMA. Usually a least-squares best fit routine is used to superpose the two experimental structures on the minimized structure.

4.4.1 Principal Component Analysis

The molecular dynamics simulations of the biological molecules are very hard to analyze in terms of identifying the motions within the molecules or functional motions. Principal component analysis (PCA) [58, 8, 99] is an excellent tool to solve this problem. Similar to NMA, PCA also incorporates the assumption that the major collective modes of fluctuation dominate the functional dynamics. The vast majority of protein dynamics can be described by a few number of collective degrees of freedom [8]. By using PCA, main modes of the collective motion can be filtered from the complex local fluctuations. The dynamics along each mode of motion can be separately analyzed and visualized. Since the principal modes of motion can usually represent the protein function, the dynamics in the low-dimensional subspace spanned by the main modes are called essential dynamics [8], meaning that principal modes are essential for the function of the molecule. Similarly, the subspace representing the major modes of

collective fluctuations is called the essential subspace. A small set of all the number of degrees of freedom of the system dominates the molecular dynamics in a biomolecule. This simplifies the analysis and interpretation of molecular dynamics trajectories. Moreover, enhanced sampling algorithms that search the essential subspace in either a systematic or exploratory fashion can be developed on this basis [62, 9, 38, 37].

In NMA of a molecular dynamics simulation, the potential is assumed to be harmonic which is not the case in PCA. Thus, PCA can be used to study the degree of anharmonicity in the molecular dynamics of a simulated system. In proteins at physiological temperatures, the major modes of collective fluctuation are dominated by anharmonic fluctuations [8, 69] and the protein dynamics has been described as diffusion among multiple minima [98, 7, 97]. On short timescales, the dynamics are dominated by fluctuations within a local minimum; this local minimum can be finely approximated by the local normal modes. On longer timescales, the large fluctuations are dominated by a largely anharmonic diffusion between multiple wells [68].

As mentioned earlier, in NMA the modes representing the largest fluctuation have the lowest frequencies. Also in PCA, the largest-amplitude modes usually represent the slowest dynamical transitions. Note that in PCA, no assumptions are implied regarding the harmonicity of the motion and the modes are usually sorted according to variance instead of frequency.

Theory

To perform a PCA, a matrix based on the fluctuations in position must be formed. It is obtained by a superposition of positions to a common reference structure and creating variance covariance matrix of positional fluctuations:

$$C = \left\langle (x(t) - \langle x \rangle)(x(t) - \langle x \rangle)^T \right\rangle \tag{4.10}$$

where, $<>$ denotes ensemble averaging. In the above equation, the coordinates x are demonstrated as a function of time but they can exist in any order. For instance, they can represent a molecular dynamics trajectory or a set of experimental structures. C is a symmetric matrix that can be diagonalized by an orthogonal coordinate transformation T:

$$C = T\Lambda T^T \tag{4.11}$$

where, Λ is the diagonal (eigenvalue) matrix and T contains the eigenvectors of C in each column. The eigenvalues λ correspond to the mean square coordinate fluctuations; therefore, each of them has a contribution to all the principal components of the total fluctuation. The eigenvectors are usually sorted such that their eigenvalues are in decreasing order. Note that

for a system of N atoms, the matrix C has $3N \times 3N$ elements. At least $3N$ configurations must be introduced in constructing the C matrix to generate $3N - 6$ eigenvectors with nonzero eigenvalues. Six of the eigenvalues with their corresponding eigenvectors are related to the overall rotation and translation of the molecule must be exactly zero. In the case where only M configurations are available (where $M < 3N$), at most $M - 1$ nonzero eigenvalues with corresponding eigenvectors will be obtained. With μ_i being the ith eigenvector of C (i.e. the ith column of T), the projection of the original configurations onto each of the principal components helps in obtaining the principal coordinates $p_i(t)$:

$$p_i(t) = \mu_i \cdot (x(t)- <x>).\tag{4.12}$$

It is worth to mention that the variance $< p_i^2 >$ is equal to the eigenvalue λ_i. For visualization purposes, the above projection can be transformed back to the Cartesian coordinates:

$$x_i'(t) = p(t) \cdot \mu_i + <x>.\tag{4.13}$$

To compare two sets of eigenvectors, μ and v, their inner products can be calculated:

$$I_{ij} = \mu_i \cdot v_j.\tag{4.14}$$

Subspace overlaps are usually calculated by summation of the squared inner products:

$$O_n^m = \sum_{i=1}^{n} \sum_{j=1}^{m} (\mu_i \cdot v_j)^2.\tag{4.15}$$

O_n^m is a measure that depicts the amount of the n-dimensional subspace of set μ which is contained within the m-dimensional subspace of set v. To achieve full overlap (O = 1), m should be larger than n [68].

4.4.2 PCA of Structural Ensembles

A PCA or essential dynamics analysis may be performed for a molecular dynamics trajectory or any other structural ensemble. The PCA consists of three main steps. The first step is to superimpose the configurations from the ensemble. This is necessary to filter the internal motions from overall rotation and translation of the molecule. The filtering is usually carried out by a least-squares fit of each of the configurations onto a reference structure. Next, a variance-covariance matrix is constructed using the fitted trajectory. The resulting matrix is symmetric;

the off-diagonal elements hold the covariances of the atomic displacements relative to their respective averages for each pair of atoms. The variances of each atom displacements are located along the diagonal of the matrix. If the atoms move together in the same direction, the positive covariances reach higher values. In contrast, anti-correlated motions result in largest negative covariances. The displacements with no correlation have near-zero covariances. Thereafter the variance-covariance matrix is diagonalized. Diagonalization of this matrix results in a set of eigenvectors and eigenvalues that must be sorted such that the eigenvalues are in decreasing order. Note that the eigenvalues represent the variance along each of the corresponding collective modes (eigenvectors). A small number of modes can be sufficient to describe the main constituents of the total fluctuation. At the final step, the original trajectory is decomposed with respect to the principal components. The trajectory must be projected onto each of the principal modes in order analyze the time behavior and distribution of each of the principal coordinates. The projections provide a representation of the sampled distribution in configuration space. They can also be helpful in comparing multiple ensembles along the principal modes of collective fluctuation. To visualize the motion along the principal coordinates, the projections onto principal coordinates must be translated back into Cartesian space.

In contrast to standard NMA, a PCA can be carried out on any subset of atoms. For the proteins, usually only C-alpha or backbone atoms are taken into account. This approximation is very critical in simulating large proteins with a number of atoms. By taking only the C-alpha or backbone atoms the required memory space and processor power in order to diagonalize the covariance matrix is significantly reduced while the main collective modes remain very similar to an all-atom analysis [8, 156]. Moreover, the backbone-only analysis will automatically exclude the artificial apparent correlations between slow side-chain fluctuations and backbone motions.

An all-atom PCA can also be compared with results from a standard NMA. In this type of analysis (a.k.a quasi-harmonic analysis), the fluctuations in PCA must be calculated from mass-weighted displacements [66]. In all-atom analysis an approximation can be employed to limit the computations to calculation of only the principal modes of fluctuation. This eliminates the need to store and diagonalize the full matrix [156]. The PCA technique has many other applications aside from the analysis of molecular dynamics trajectories. For example, PCA can be used to derive the principal modes from sets of x-ray structures [157] or to compare simulation data with experimental conformations [43, 40, 3]. PCA is also helpful in determining search directions from multiple homologous structures in homology modeling [133].

Convergence of PCA Results Derived from Molecular Dynamics Simulations

One can use the principal components obtained from different simulations or simulation parts to compare the major directions of configurational space and sampled regions which aids in

analyzing similarity and convergence. The sub-nanosecond molecular dynamics simulations for proteins are known to have a significant sampling problem which causes a poor overlap between the principal components extracted from multiple parts of these trajectories [16, 31]. Although the individual principal components may not be the same, the subspaces derived from the major principal components converge rapidly. They are also consistent within different simulation results and between simulations and experiments [43, 3, 6, 39]

Principal coordinates with high cosine content usually indicate a nonconverged trajectory. But the lack of convergence of the dynamics along a set of modes does not mean that the directions of such modes or the subspace they span are not converged. When an acceptable converged trajectory is achieved, the thermodynamic properties can be calculated in form of ensemble averages. They can be mapped onto the principal coordinates to visualize the properties such as free energy landscapes.

Chapter 5

Results

The results from classical MD simulations for Prolyl Oligopeptidase (POP) with and without Z-pro-prolinal (ZPP) are presented in this chapter. First the POP protein without ZPP inhibitor is simulated for 50ns. Afterwards, the ZPP inhibitor is put in the binding cavity of the POP, but without the hemiacetal bond; so ZPP is free to move all over the binding cavity.

The POP protein structure is studied and properties such as radius of gyration, root mean square deviation (RMSD) and fluctuations of root mean square deviation (RMSFs) are calculated and compared with the previous simulations [95] where ZPP inhibitor is in the binding cavity of the POP without the hemiacetal bond activated. Also, the simulated radius of gyration (R_g) and RMSD and RMSD fluctuation of the POP protein with ZPP inhibitor and without ZPP inhibitor are compared. The principal component analysis (PCA) method is employed to examine dynamics of the protein molecule and its principal modes of motion. The displacements of the atoms of POP molecule are examined for an isolated POP and POP with the presence of ZPP, in order to determine whether ZPP can find a way to slip outside of the protein or not. As we explained before, for PCA of the protein (POP or POP with ZPP) C-alpha or backbone atoms are taken into account.

5.1 Prolyl Oligopeptidase without ZPP

5.1.1 Root Mean Square Deviation

Fig. 5.1 presents the time development of the root mean square displacement (RMSD) of the protein backbone atom over 50ns (the whole simulation time). The RMSD values become bounded roughly in between 0.15nm and 0.2nm after the simulation reaches a stable condition. Therefore, the simulation is considered to be stable after 25 ns.

RMSD
Backbone after lsq fit to Protein

Figure 5.1: Time development of the root mean square displacement (RMSD) of the protein backbone atom (POP without ZPP) over the whole simulation time.

Fig. 5.2 shows the RMSD values for POP with non-bonded ZPP from reference [95]. Fig. 5.1 and Fig. 5.2 demonstrate similar behaviour for both cases.

5.1.2   Radius of Gyration (R_g)

Fig. 5.3 presents time development of the radius of gyration (R_g) of the POP without ZPP over 50ns. The R_g time history indicates a stable behaviour for the whole simulation time.

Fig. 5.4 presents R_g for POP with non-bonded ZPP from reference [?]. Comparing Fig. 5.3 and Fig. 5.4 shows that both POP without ZPP and POP with non-bonded ZPP have roughly
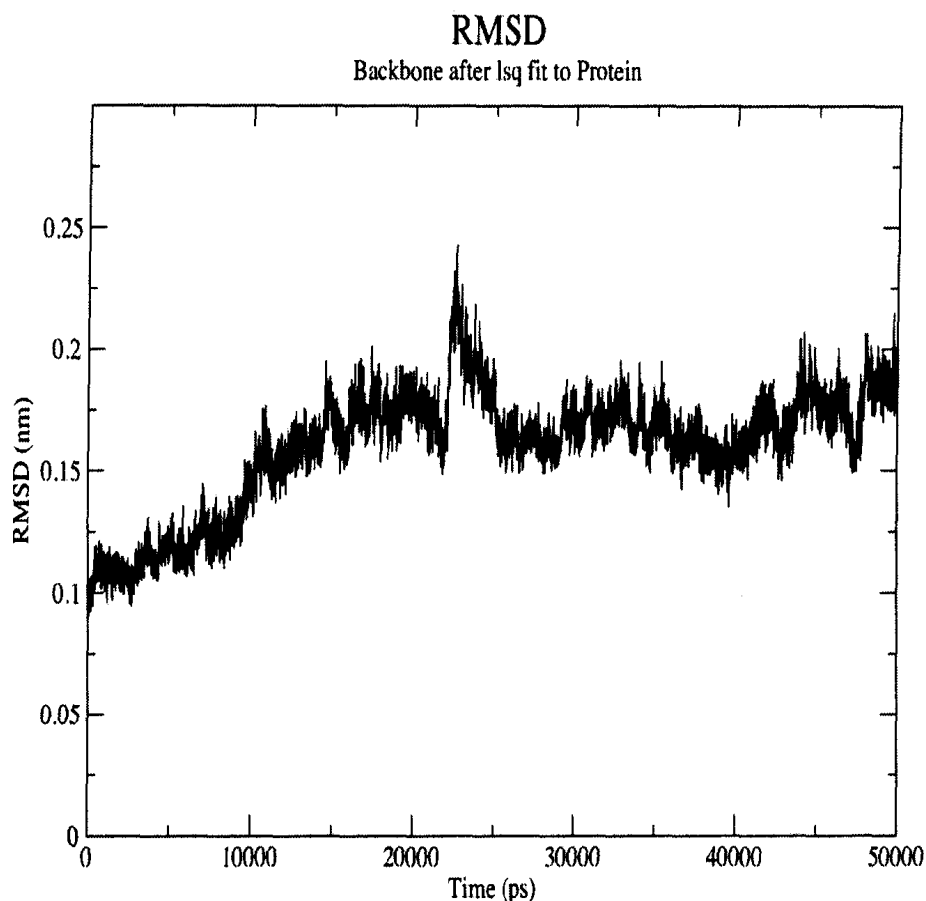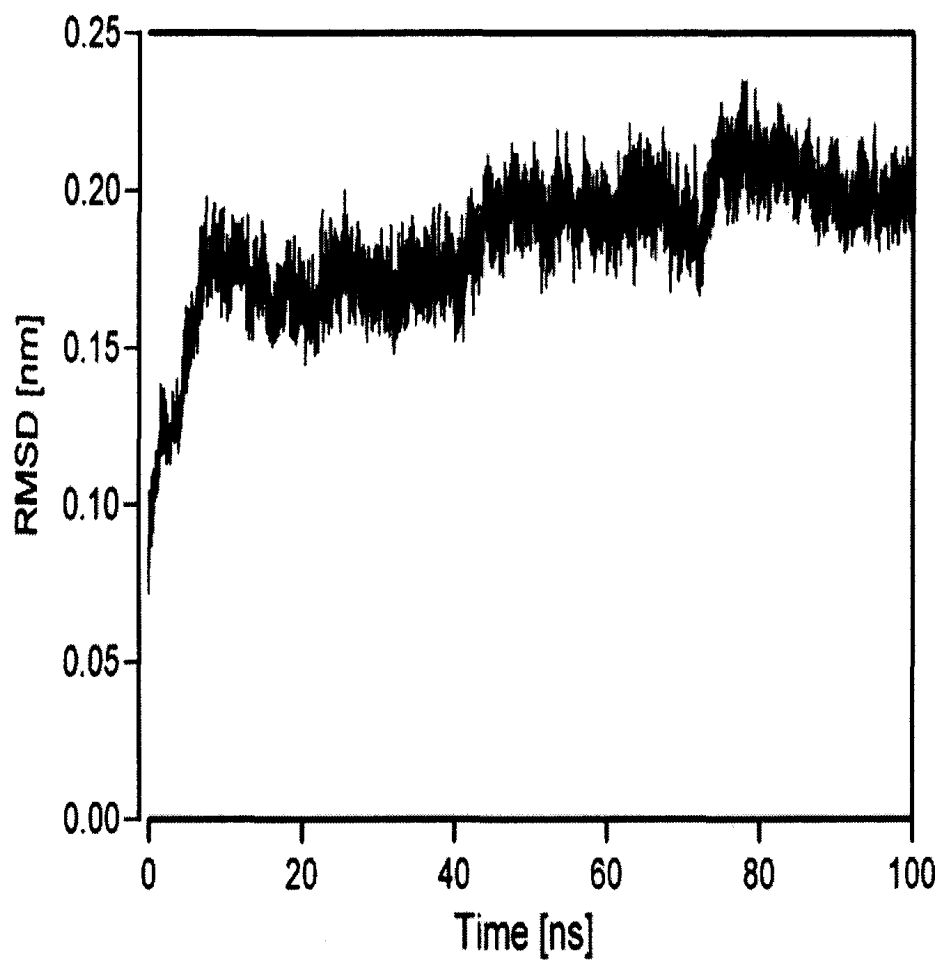
Figure 5.2: Time development of the root mean square displacement (RMSD) of the POP with non-bonded ZPP backbone atom over the whole simulation time (From [95] with permission. ©: Taylor and Francis)
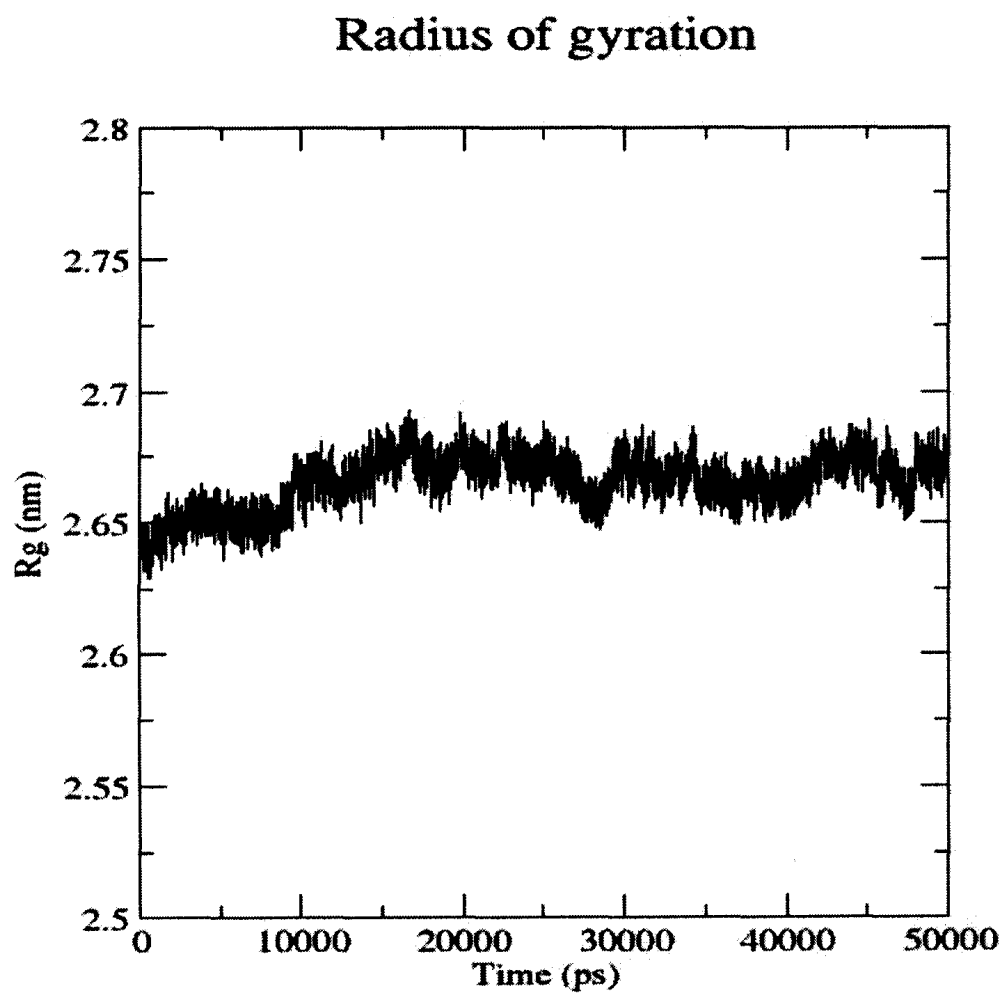
Radius of gyration

Figure 5.3: Time development of the Radius of Gyration (R_g) of the POP without ZPP over the whole simulation time.
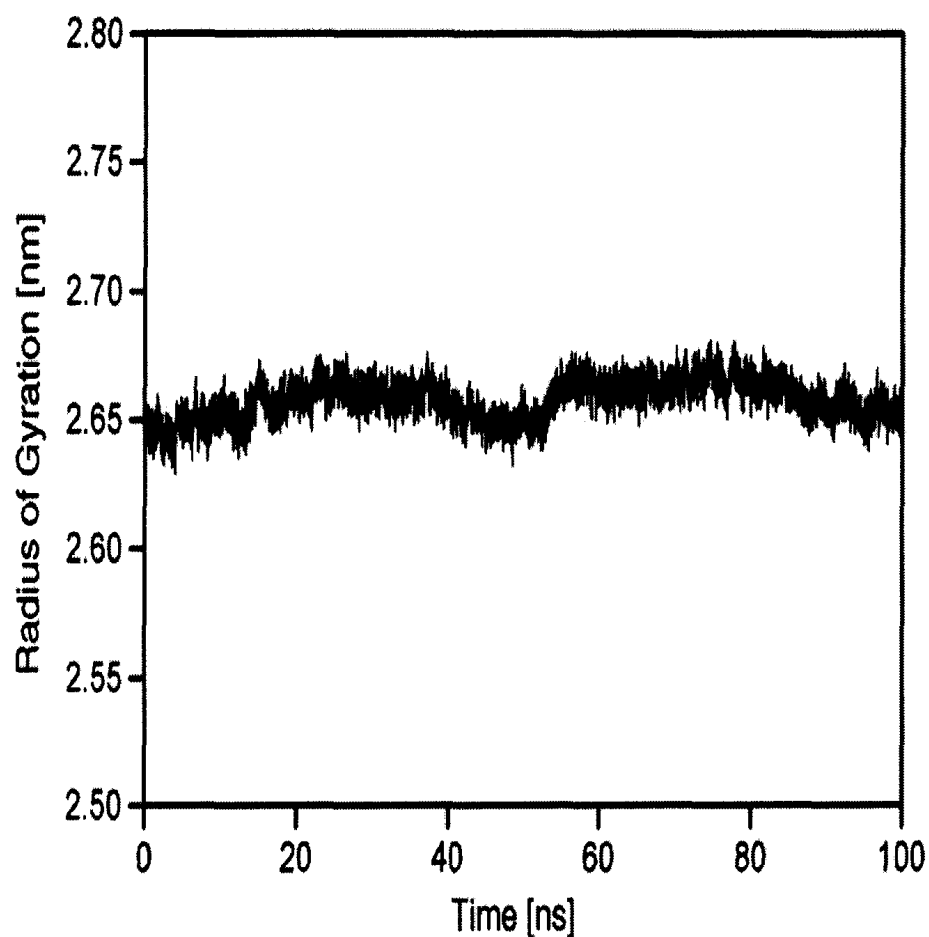
Figure 5.4: Time development of the Radius of Gyration (R_g) of the POP with non-bonded ZPP over the whole simulation time (From [95] with permission. ©: Taylor and Francis)

equal radius of gyration. Note that the observed small difference in R_g has originated from the presence of ZPP.

Radius of gyration for backbone of the POP is shown Fig. 5.5. The R_g for backbone of the molecule follows the same trend as the molecule R_g but the values are approximately 0.05nm lower than those observed for the whole molecule.

5.1.3 Root Mean Square Displacement Fluctuations

The RMSF values are calculated for the C-alpha atoms. The RMSFs are used to describe the basic dynamics of the protein. In Fig. 5.6 the lower fluctuations for the rigid structural ele-
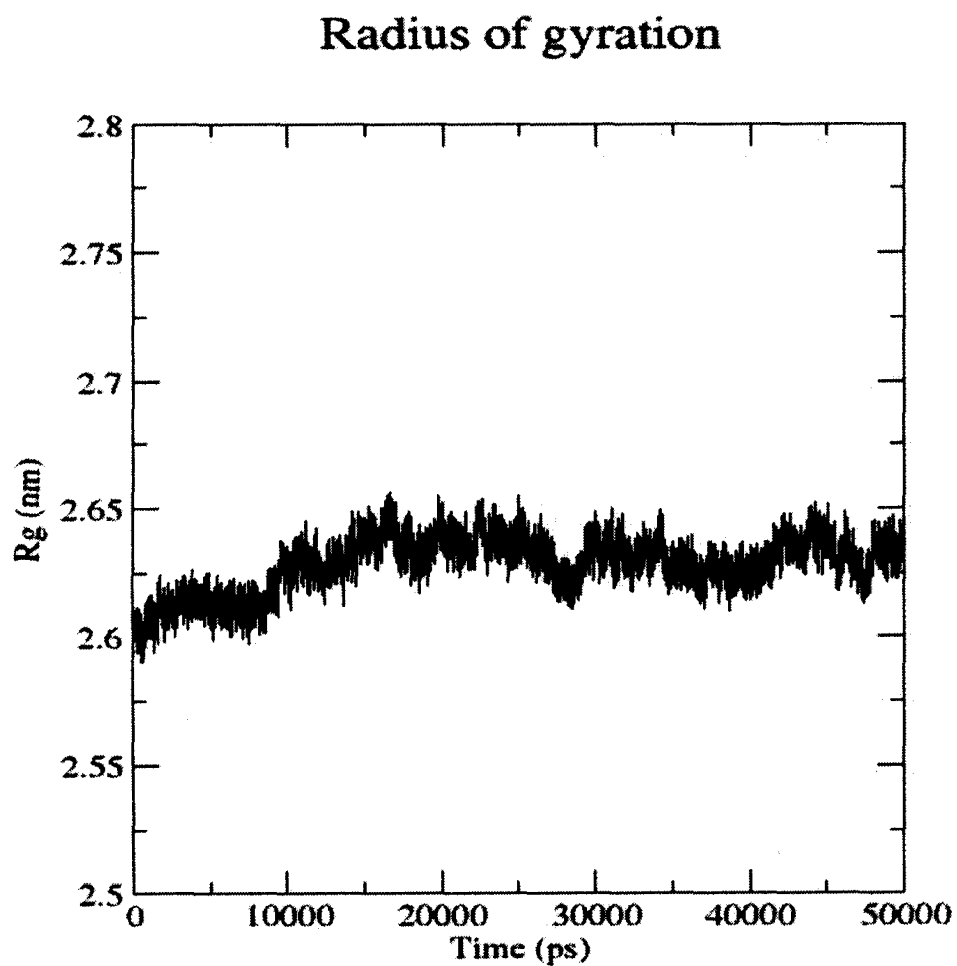
Radius of gyration

Figure 5.5: Time development of the Radius of Gyration (R_g) of the backbone over the whole simulation time for POP without ZPP.

RMS fluctuation
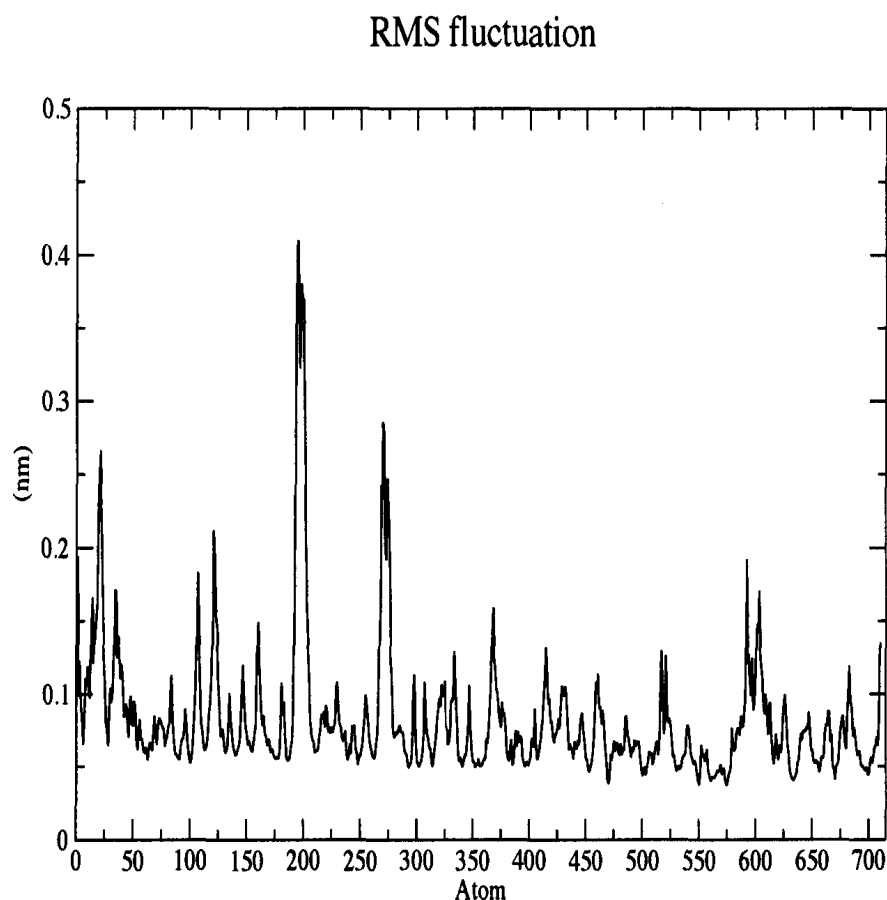
Figure 5.6: RMSFs of the C-alpha atoms for POP without ZPP corresponding to the residues measured from the entire molecular dynamics run. The RMSFs show small values for the rigid structural elements and larger values for the ends and loops. The large values are: 21GLY, 122ASP, 196LYS, 198ASP, 271ASN.

ments and larger values for the ends and loops can be easily seen; this is exactly as expected for most of the protein simulations.

Fig. 5.6 and Fig. 5.7 are slightly different from each other due to the presence of ZPP inside POP.

The larger values in Fig. 5.6 show that the loops formed by ASP (198), ASN (271), ASP (122) and GLY (21) residue are flexible. The mentioned loops and residues are shown in Fig. 5.8 and Fig. 5.9.
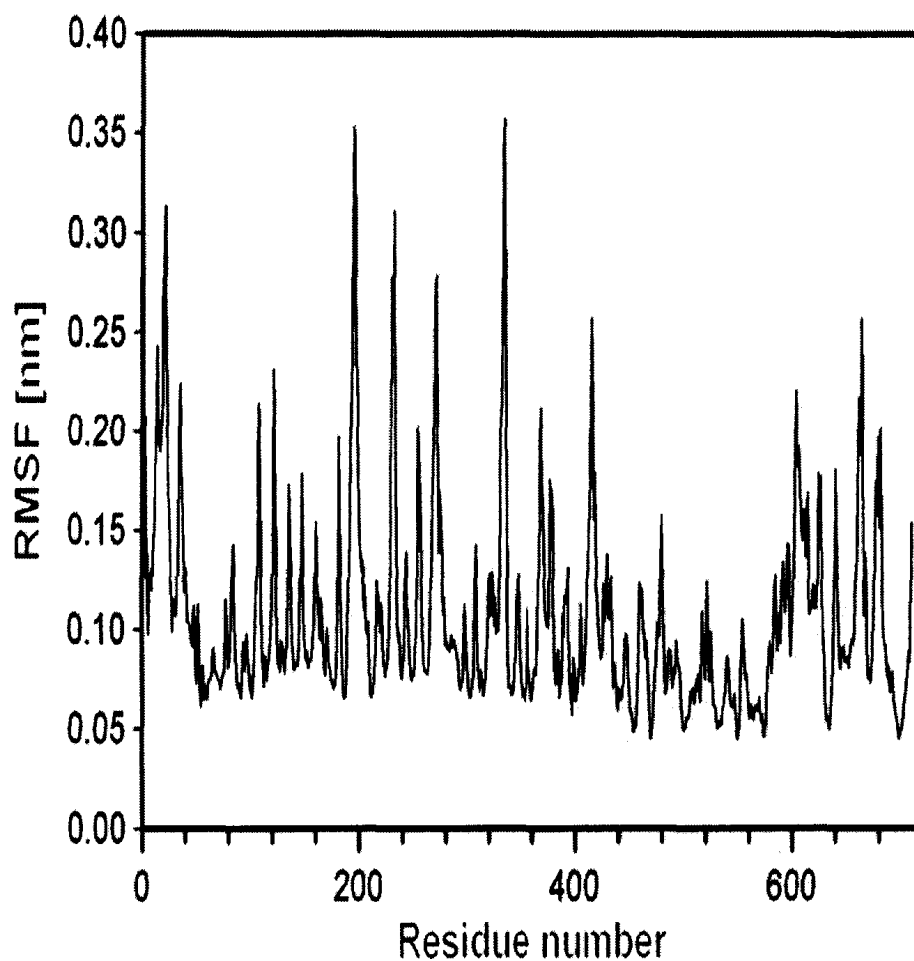
Figure 5.7: RMSFs of the C-alpha atoms for POP with non-bonded ZPP corresponding to the residues measured from the entire molecular dynamics run. The RMSFs show small values for the rigid structural elements and larger values for the ends and loops (From [95] with permission. ©: Taylor and Francis)
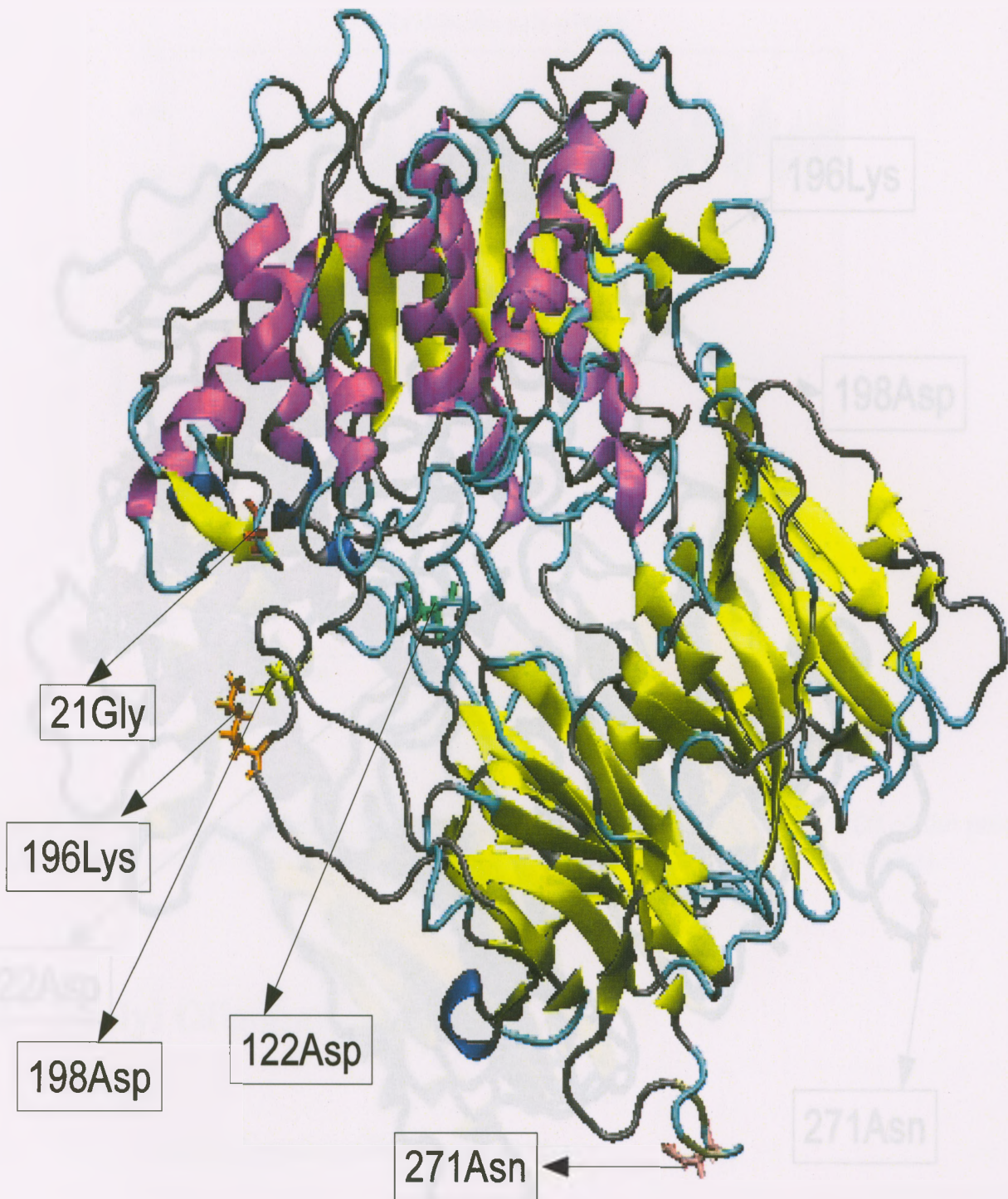
Figure 5.8: The residue number for the largest values of the RMS fluctuations for POP without ZPP. In the Fig. the ASP (198), ASN (271), ASP (122) and GLY (21) residues are labeled in their loops.
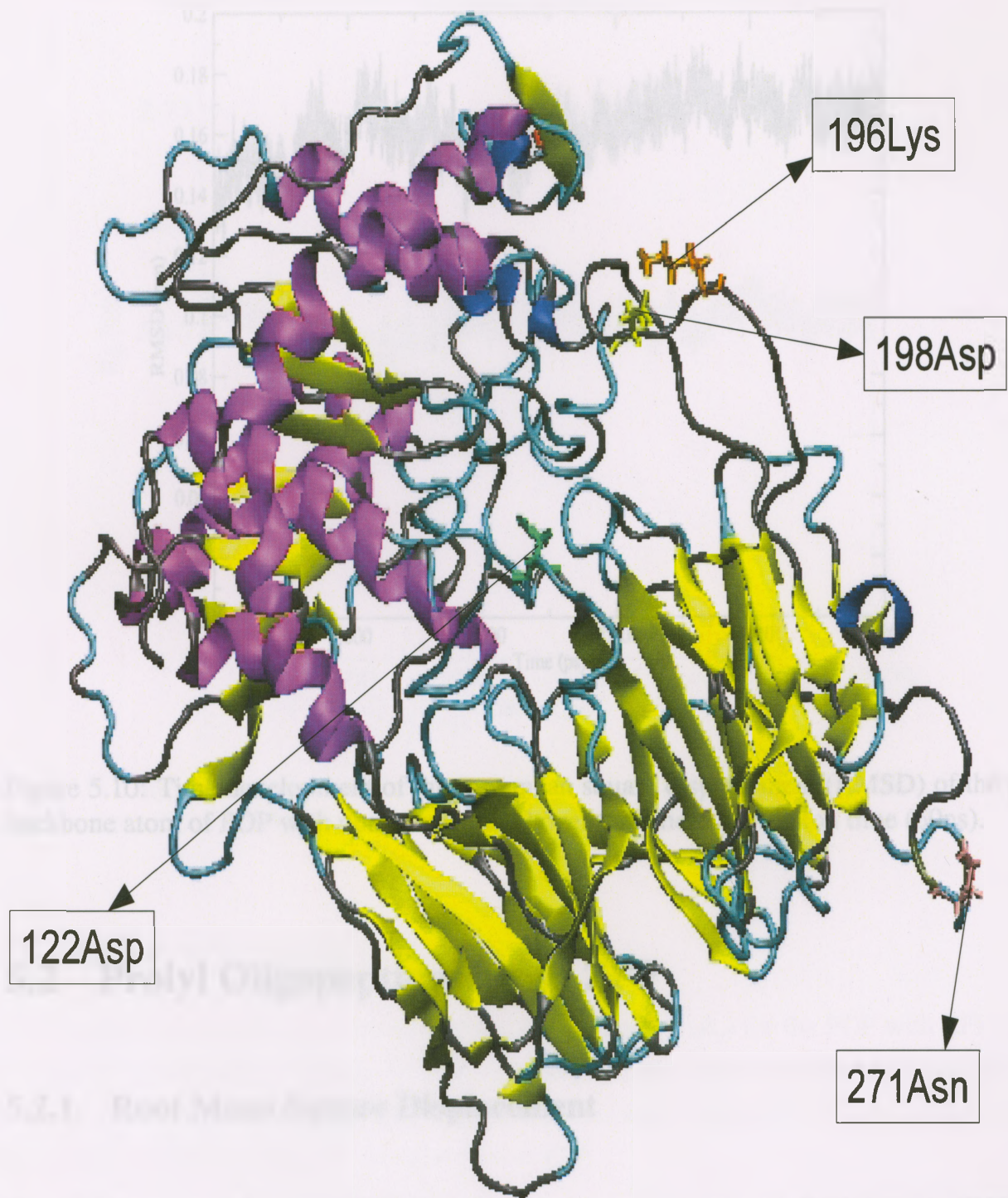
Figure 5.9: Same as Fig. 5.8 but viewed from different angle for POP without ZPP.

RMSD
Backbone after lsq fit to Protein

Figure 5.10: Time development of the root mean square displacement (RMSD) of the protein backbone atom of POP with non-bonded ZPP over the whole simulation time (50ns).

5.2 Prolyl Oligopeptidase with ZPP

5.2.1 Root Mean Square Displacement

Fig. 5.10 presents the time development of the root mean square displacement (RMSD) of the protein backbone atoms for POP with non-bonded ZPP over 50ns (the whole simulation time). The RMSD values become bounded roughly in between 0.15nm and 0.19nm after the simulation reaches a stable condition. The simulation is considered to become stable after 25 nano seconds.

Radius of gyration
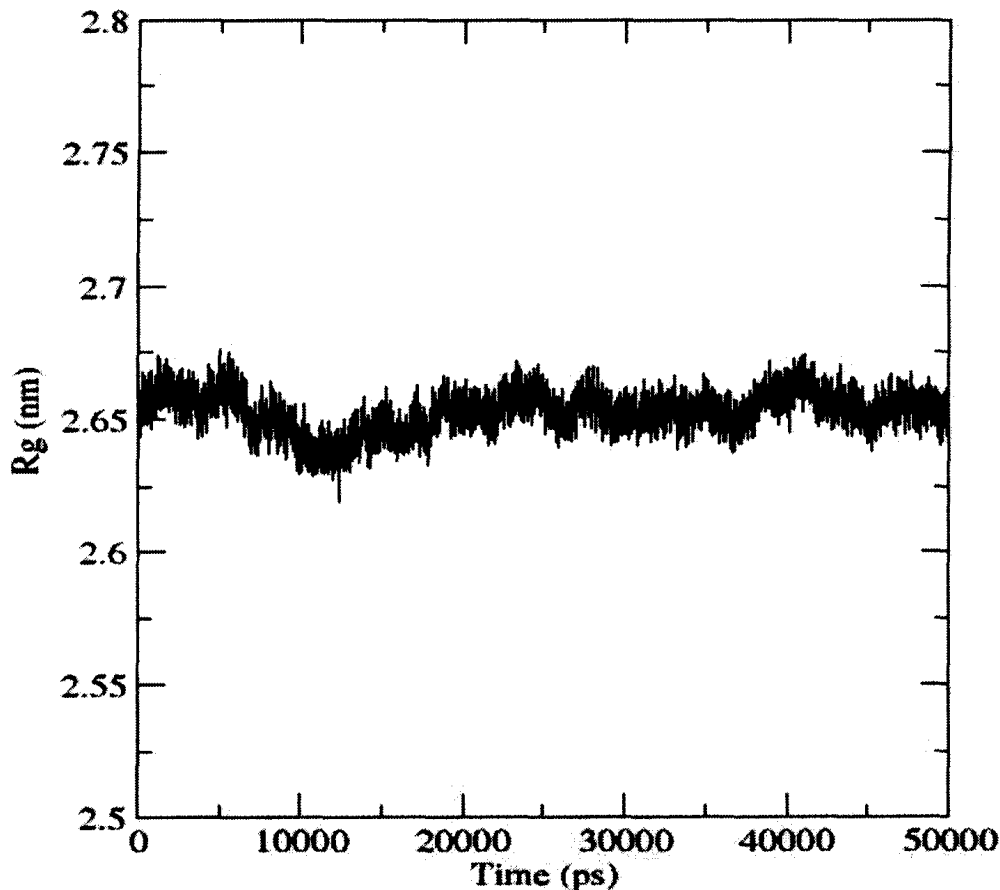
Figure 5.11: Time development of the Radius of Gyration (R_g) of the protein over the whole simulation time (50ns) for POP with non-bonded ZPP.

5.2.2 Radius of Gyration (R_g)

Fig. 5.11 presents time development of the Radius of Gyration (R_g) for the POP with ZPP over 50ns. The R_g time history indicates a stable behaviour for the whole simulation time. Fig. 5.3 and Fig. 5.11 demonstrate similar radii of gyration after approximately 20 nano seconds. The difference between them is due to the presence of ZPP.

Fig. 5.12 shows radius of gyration of the backbone atom over the whole simulation (50ns) time for POP with non-bonded ZPP. Fig. 5.11 and Fig. 5.12 have the same trend but the R_g for the backbone is lightly shifted below the R_g for the protein.

Fig. 5.13 presents the radius of gyration of POP and ZPP over the whole simulation (50ns)
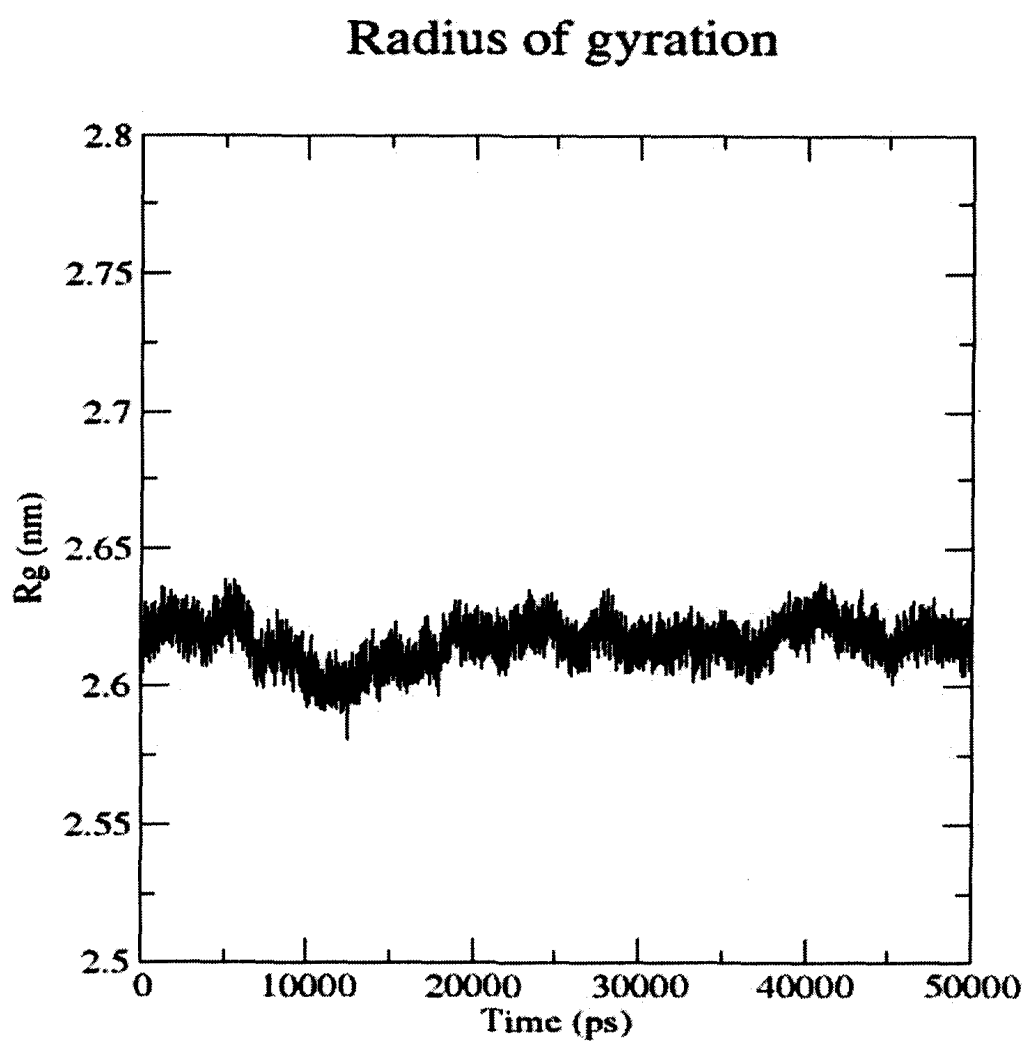
Radius of gyration

Figure 5.12: Time development of the Radius of Gyration (R_g) of the backbone atom over the whole simulation (50ns) time for POP with non-bonded ZPP.

Radius of gyration
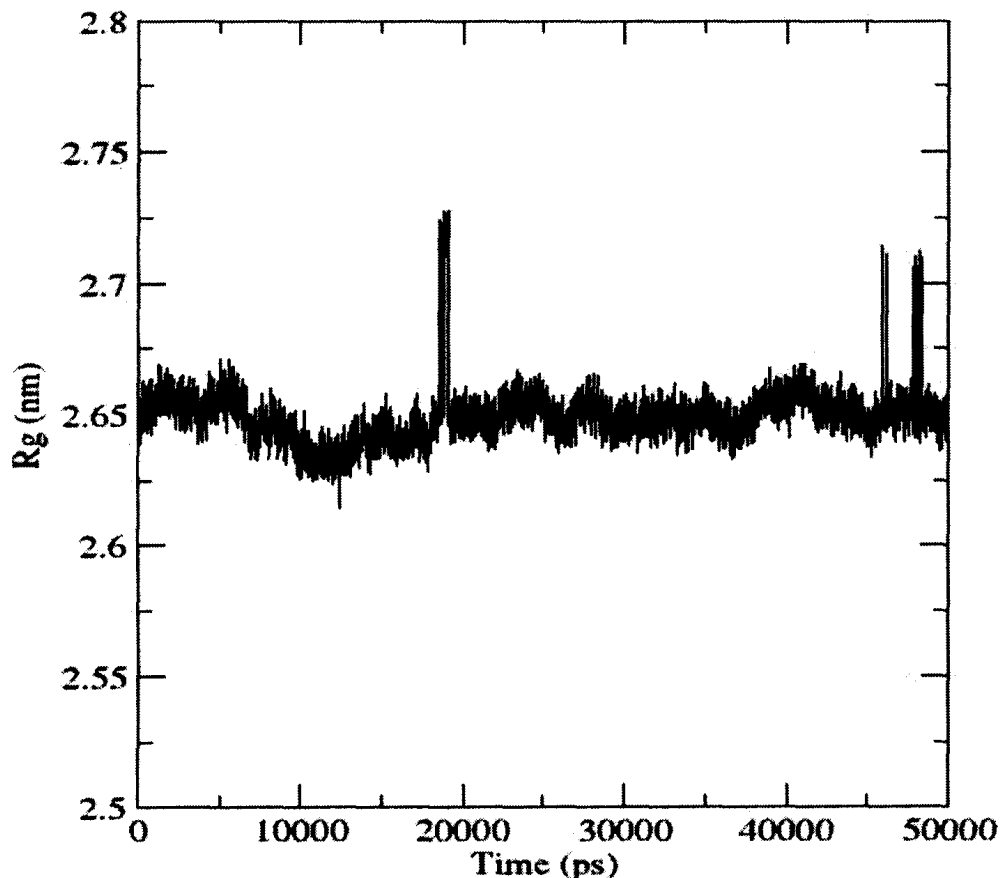
Figure 5.13: Time development of the Radius of Gyration (R_g) of the both POP and ZPP over the whole simulation time for POP with non-bonded ZPP.

time for POP with ZPP. This figure is similar to Fig. 5.11, but in Fig. 5.13 the ZPP is considered in calculations. The peaks in Fig. 5.13 have appeared in presence of ZPP.

5.2.3 Root Mean Square Displacement of Fluctuations

RMSF values are calculated for the C-alpha atoms. As previously mentioned, the RMSFs are small for the rigid structural elements and the large RMSFs belong to loops and ends. With the same rational, it can be concluded that in Fig. 5.14 larger values of RMSFs belong to 1Met and 710Pro end residues. The loops that include 21Gly, 34Pro, 58Pro, 196Lys, 230Asp, 270SER, 323Glu, 345Arg, 368Thr, 415Glu, 462Leu, 603Asp, 494His and 686Val residues have the greatest values. The mentioned residues are labeled in Fig. 5.15 and Fig. 5.16.

5.3 Principal Component Analysis

5.3.1 POP without ZPP

Fig. 5.17 shows that only a few of the eigenvalues of the covariance matrix have large values. The first few large eigenvalues are important in analyzing the protein. These eigenvalues correspond to the main modes of motions in the proteins. As it is clear, only the first few eigenvalues are large, and the eigenvalues after 5th one can be neglected which also implies the first few modes of motion are important.

Fig. 5.18 (RMSFs of the C-alphas for first eigenvector of POP without ZPP) confirms that residues 196LYS and 271ASN are more flexible as demonstrated in the Fig. 5.6 (RMSFs of the C-alphas for POP without ZPP). These residues are labeled in Fig. 5.8 and Fig. 5.9.

Fig. 5.18 shows the first eigenvector of POP. It is observed that 196Lys, 198Asp and 271Asn have the largest movements. This can also be seen in Fig. 5.19 by examining the length of the arrows attached to the residues.

Fig. 5.20 presents the RMSF vaules of the second eigenvector corresponding to C-alpha atoms in POP without ZPP. The peaks which are larger than 0.1nm in the graph belong to the motions of the residues 21Gly, 196Lys, 198Asp and 271Asn. The motion of these residues are shown in Fig. 5.21 and Fig. 5.22.

5.3.2 POP with ZPP

Eigenvectors for POP with free ZPP inside cavity are arranged from highest to lowest values and shown in Fig. 5.23. As expected, only the first few eigenvalues are large, and the eigenvalues after 5th one can be neglected. This means only the first few modes of motion are important.

Fig. 5.24 shows RMSFs of the C-alpha atoms for first eigenvector of POP with free ZPP. The peaks larger than 0.1nm in this graph belong to the ends and loops containing 1Met, 21Gly, 34Pro, 230Asp, 270Ser, 323Glu, 345Agr, 603Leu, 707Asp and 710Pro residues. These residues are labeled in Fig. 5.25. Note that the ZPP atom is not shown to maintain the clarity

RMS fluctuation
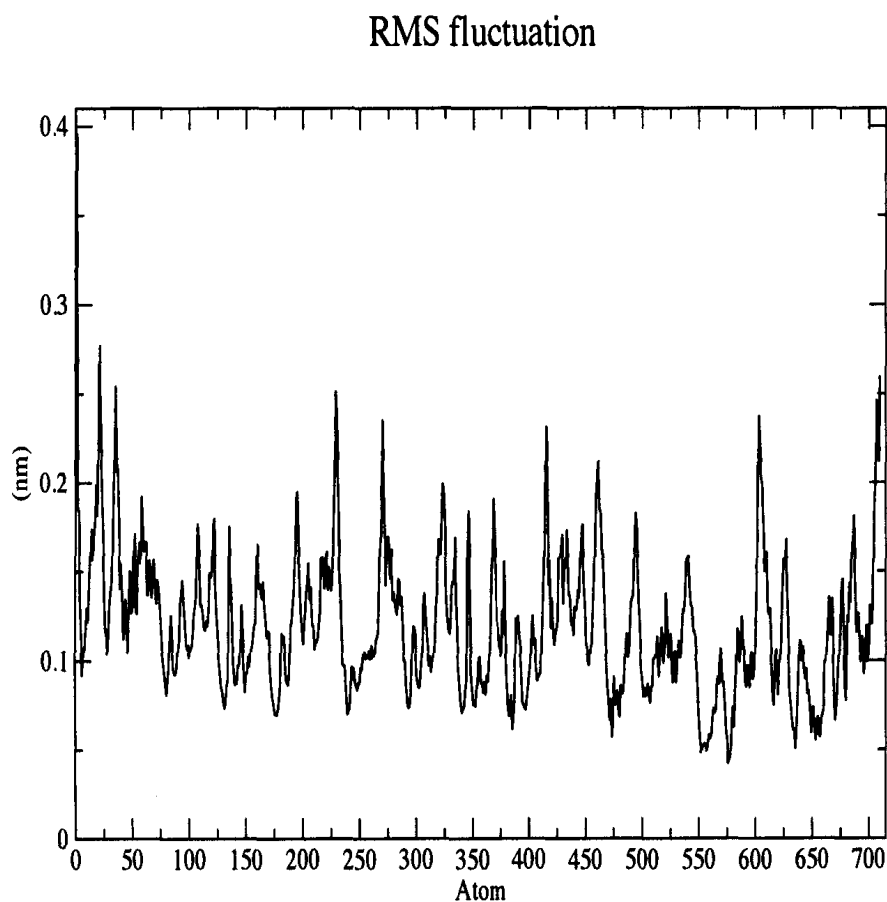
Figure 5.14: RMSFs of the C-alpha atoms for POP with non-bonded ZPP corresponding to the residues measured from the entire molecular dynamics run. The RMSFs show small values for the rigid structural elements and larger values for the ends and loops. The large values of the RMS fluctuation are at ends and loops. The loops that include 21Gly, 34Pro, 58Pro, 196Lys, 230Asp, 270SER, 323Glu, 345Arg, 368Thr, 415Glu, 462Leu, 603Asp, 494His and 686Val residues have the greatest values. These residues are shown in next two figure (in different viewing angle).
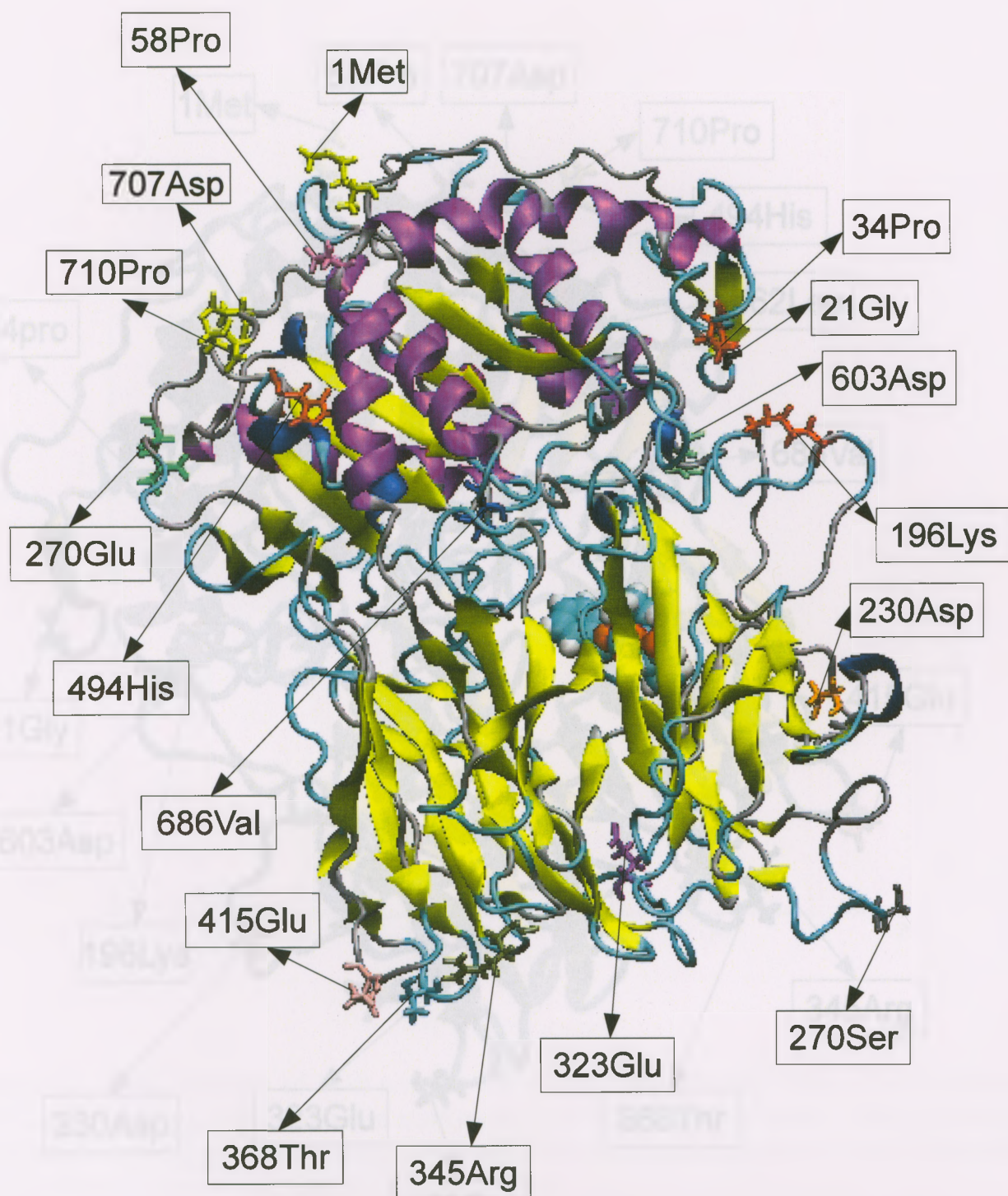
Figure 5.15: The residues with larger values of the RMSD fluctuation which appeared as peake values in Fig. 5.14 for POP with non-bonded ZPP. The loops that include 21Gly, 34Pro, 58Pro, 196Lys, 230Asp, 270SER, 323Glu, 345Arg, 368Thr, 415Glu, 462Leu, 603Asp, 494His and 686Val residues have the greatest RMSD values.
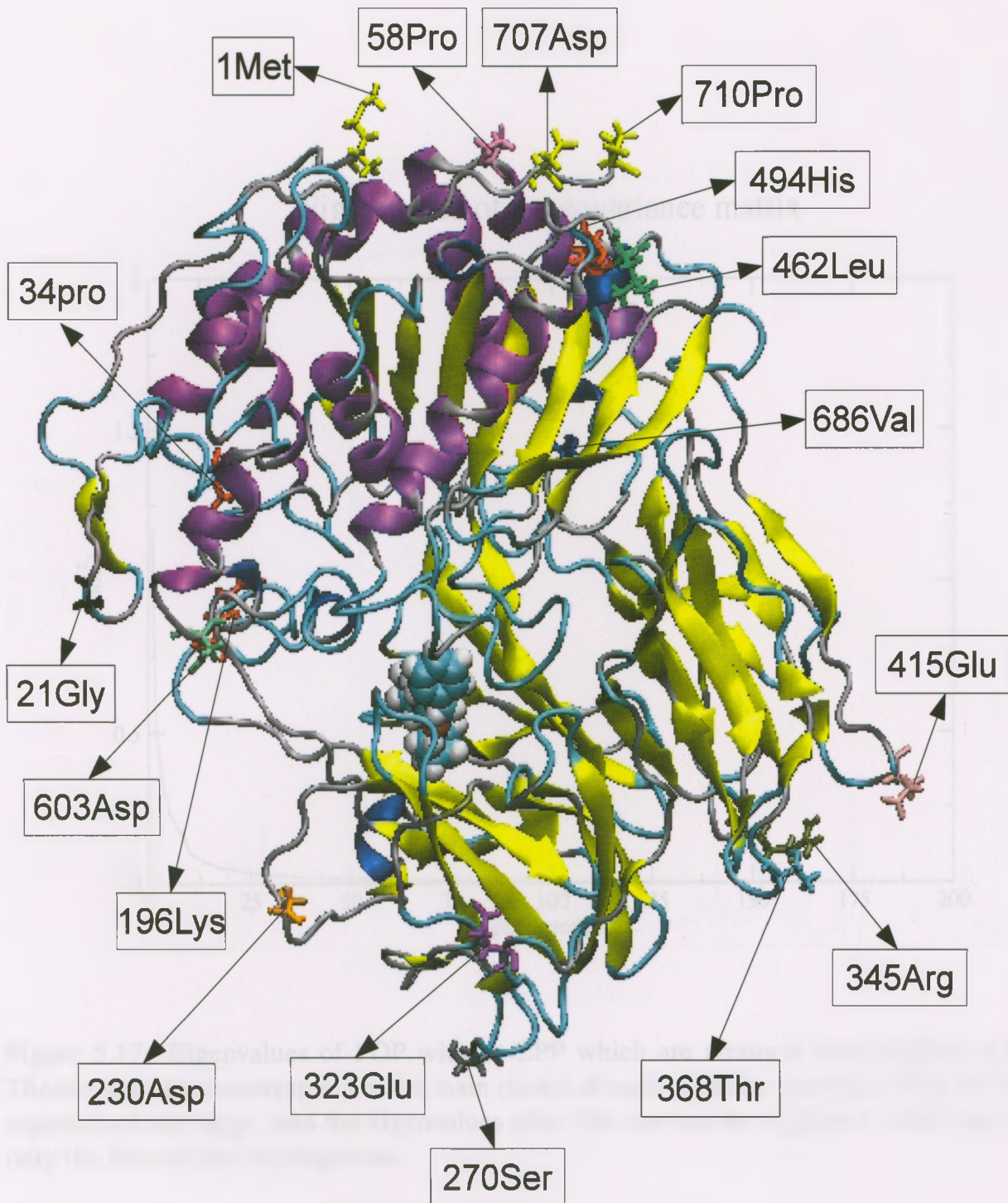
Figure 5.16: Same as Fig. 5.15 but viewed from a different angle for POP with non-bonded ZPP.

Eigenvalues of the covariance matrix
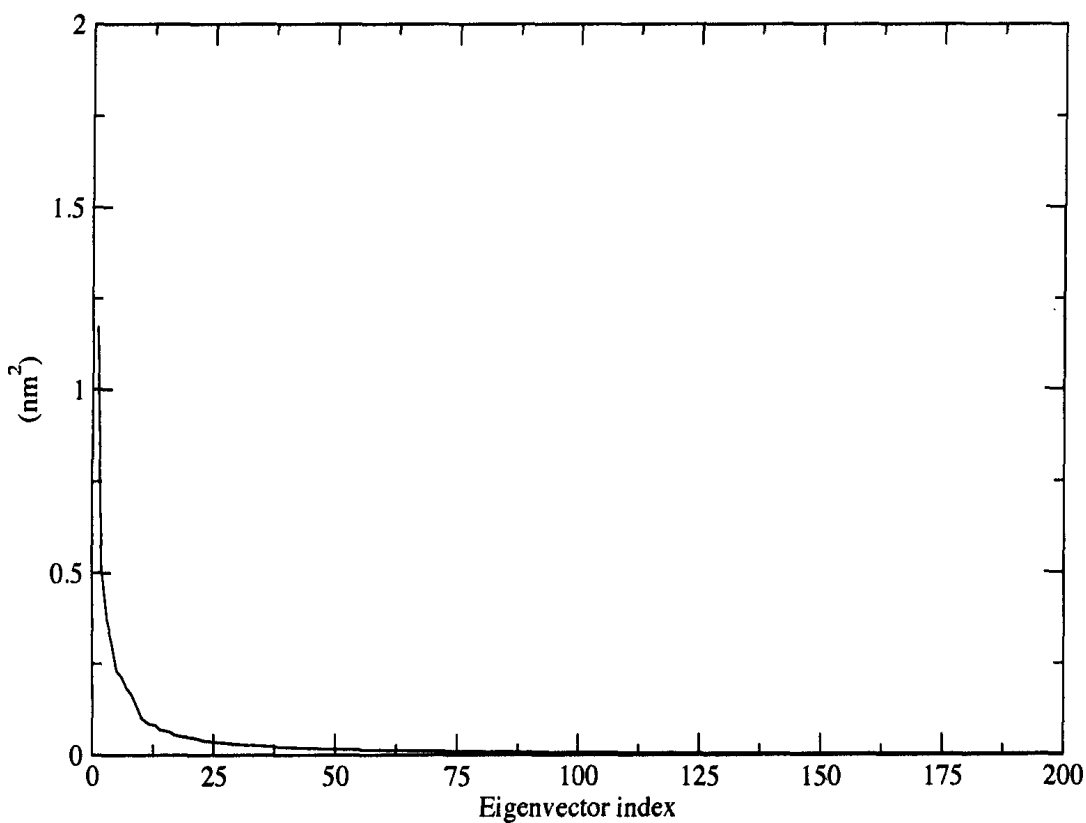
Figure 5.17: Eigenvalues of POP without ZPP which are arranged from highest to lowest. These eigenvalues correspond to the main modes of motions in the proteins. Only the first few eigenvalues are large, and the eigenvalues after 5th one can be neglected which also means only the first modes are important.

RMS fluctuation (nm)

Figure 5.18: RMSFs of the C-alpha atoms for first eigenvector of POP without ZPP corresponding to the residues measured from the entire molecular dynamics run. The RMSFs have small values for the rigid structural elements and larger values for the ends and loops. The peaks relate to residues 196LYS and 271ASN which have larger movements.
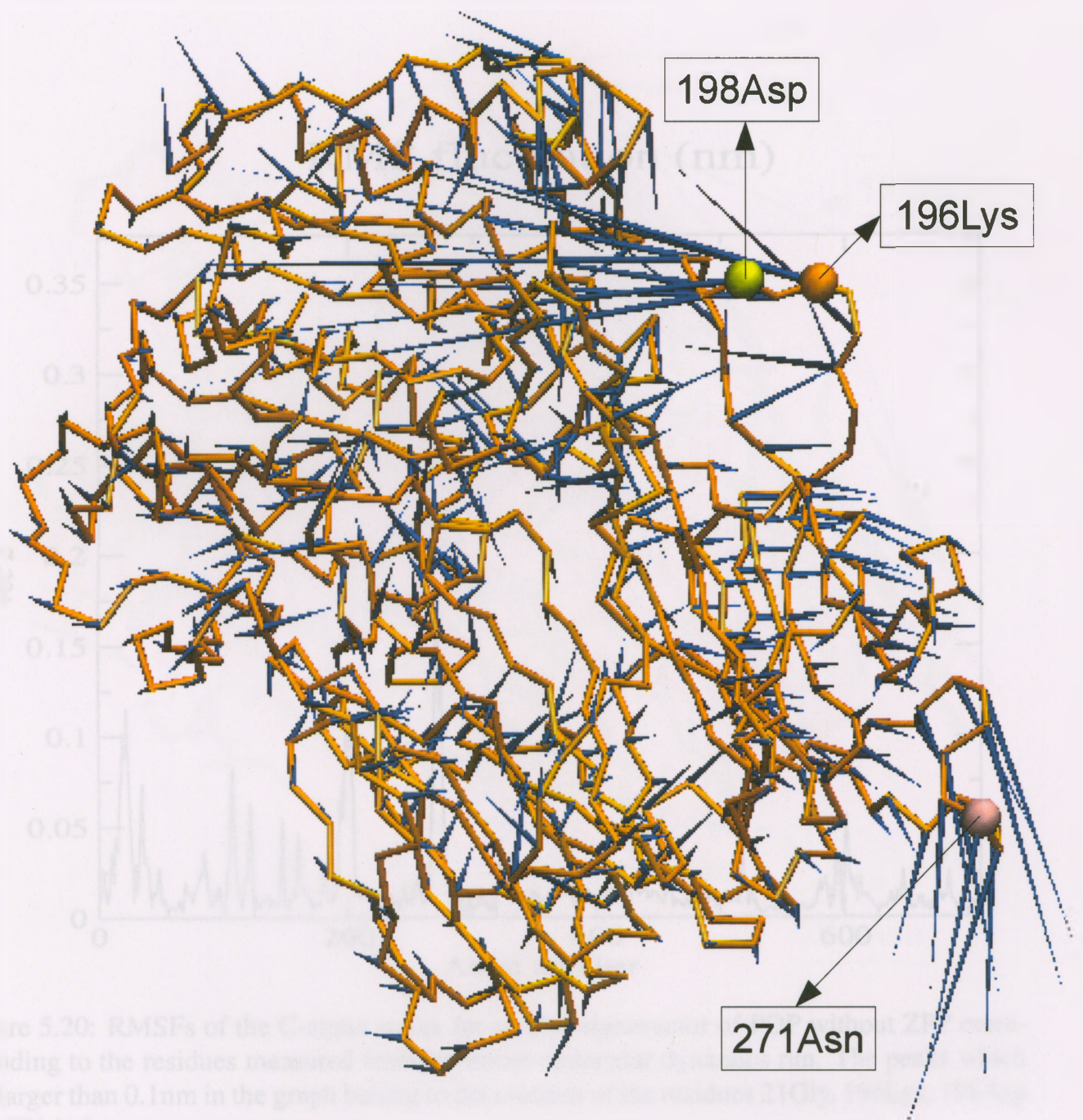
Figure 5.19: The loops containing residues 196LYS, 198ASP and 271ASN have more displacements compared to the rest of the loops. They can in fact open and close and control the access to cavity- POP without ZPP, from first eigenvector for C-alpha atoms.

Figure 5.20: RMSFs of the C-alpha atoms for second eigenvector of POP without ZPP corresponding to the residues measured from the entire molecular dynamics run. The peaks which are larger than 0.1nm in the graph belong to the motions of the residues 21Gly, 196Lys, 198Asp and 271Asn.
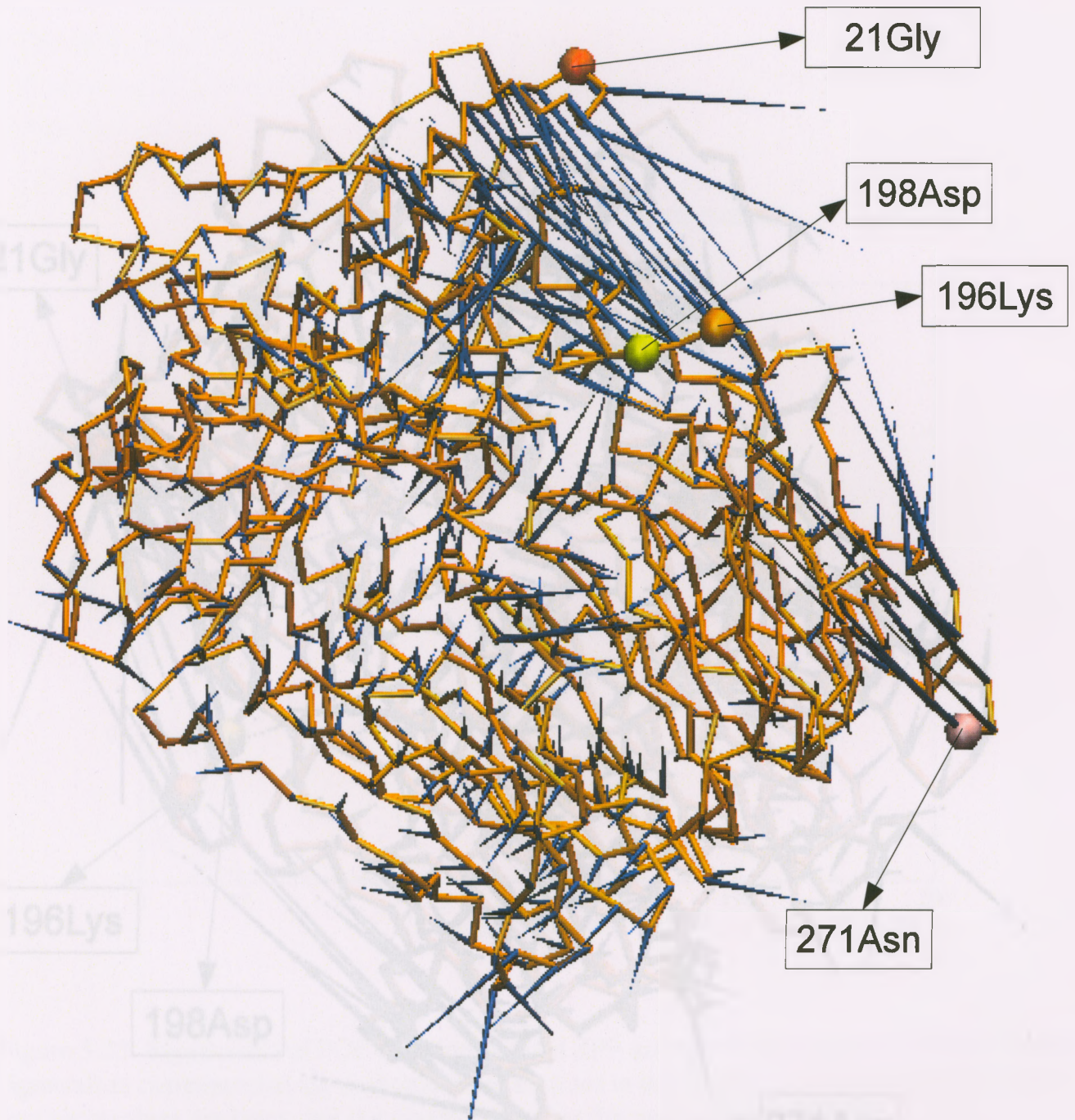
Figure 5.21: The loops containing the residues 271ASN,196LYS, 198ASP and 21GLY have longer movements compared to the rest of loops. Therefore, They can open and close the opening to the protein cavity (also shown in Fig. 5.8 and Fig. 5.9)- POP without ZPP, second eigenvector, C-alpha atoms.
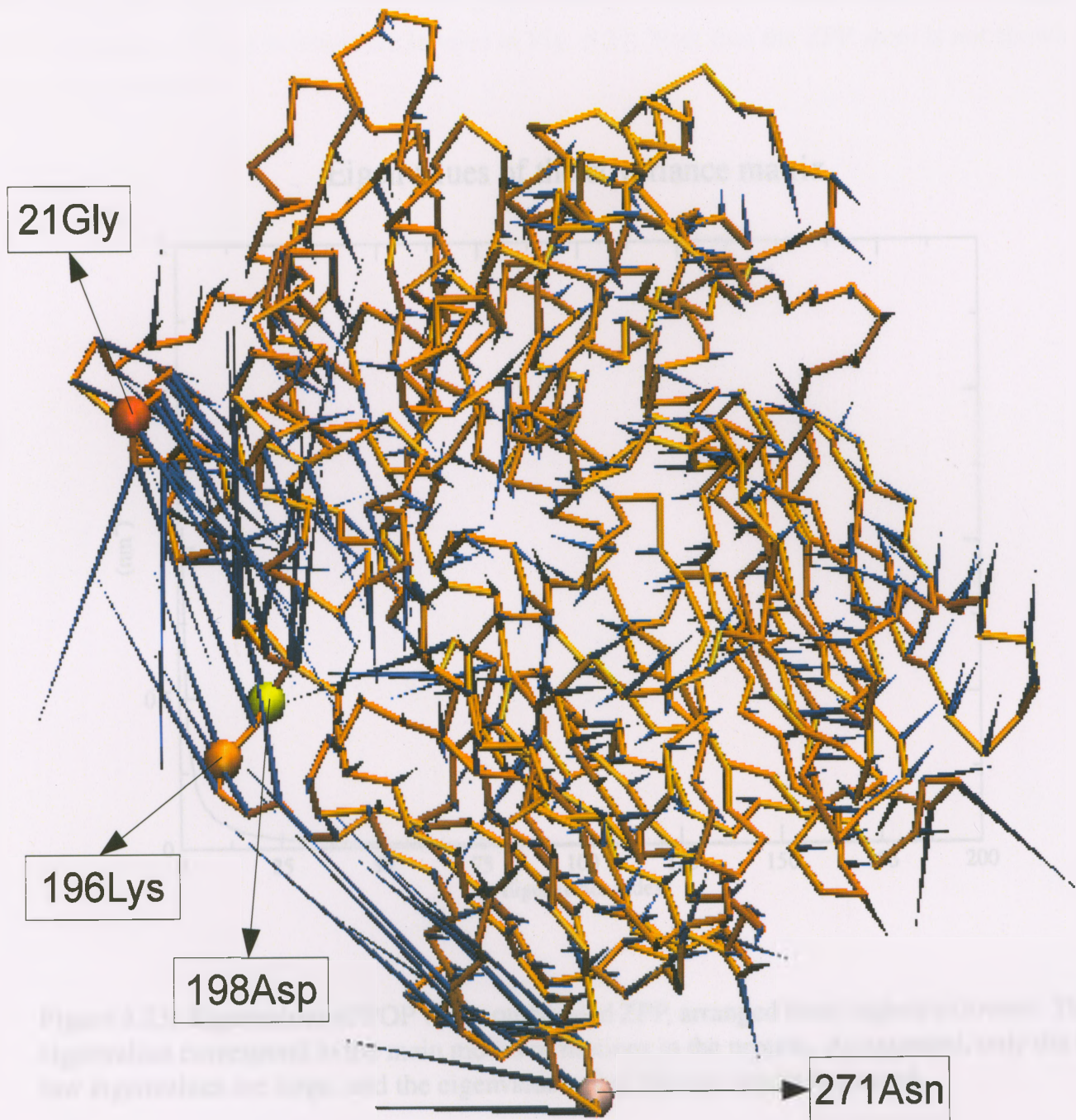
Figure 5.22: Same as Fig. 5.21 but viewed from a different angle.
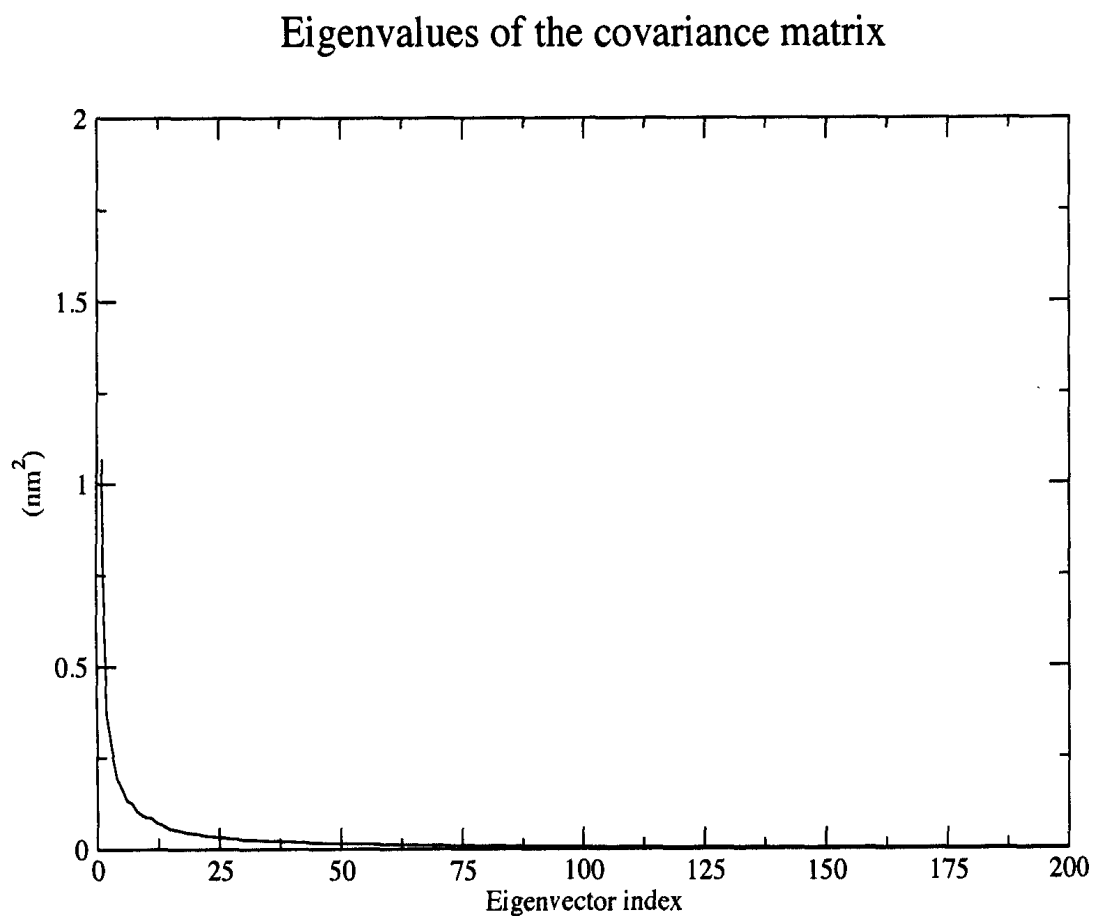
Figure 5.23: Eigenvalues of POP with non-bonded ZPP, arranged from highest to lowest. These eigenvalues correspond to the main modes of motions in the protein. As expected, only the first few eigenvalues are large, and the eigenvalues after 5th one can be neglected.

of the figure.

Fig. 5.26 shows RMSFs of the C-alpha atoms for the second eigenvector of POP with free
ZPP. The peaks larger than 0.75nm in this graph belong to the loops containing 34Pro, 415Glu,
686Val residues. These residues are labeled in Fig. 5.27. Note that the ZPP atom is not shown
in the demonstration.

RMS fluctuation (nm)
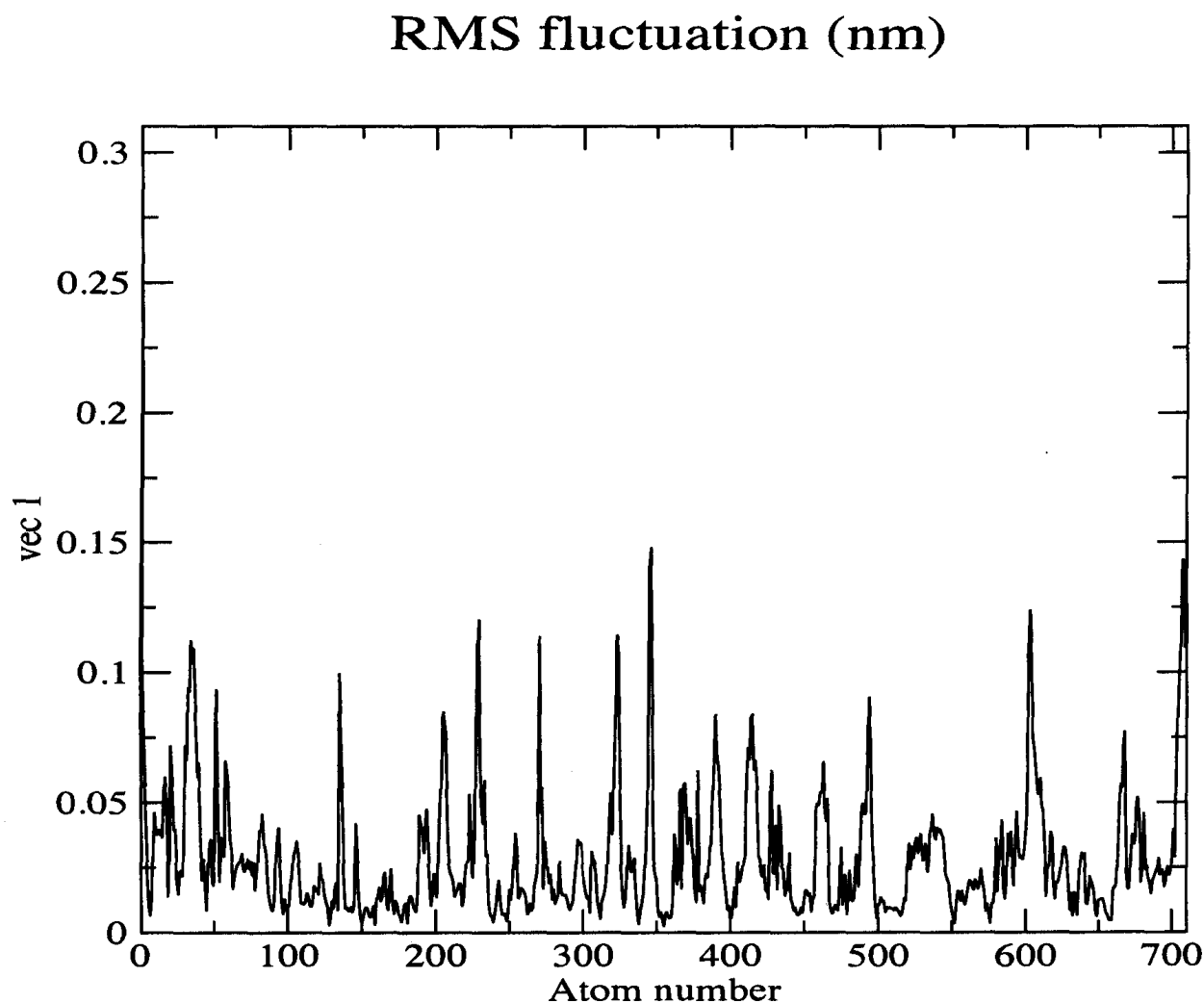
Figure 5.24: RMSFs of the C-alpha atoms for first eigenvector of POP with free ZPP corresponding to the residues measured from the entire molecular dynamics run. The RMSFs have small values for the rigid structural elements and larger values for the ends and loops. The peaks larger than 0.1nm in this graph belong to the ends and loops containing 1Met, 21Gly, 34Pro, 230Asp, 270Ser, 323Glu, 345Agr, 603Leu, 707Asp and 710Pro residues.
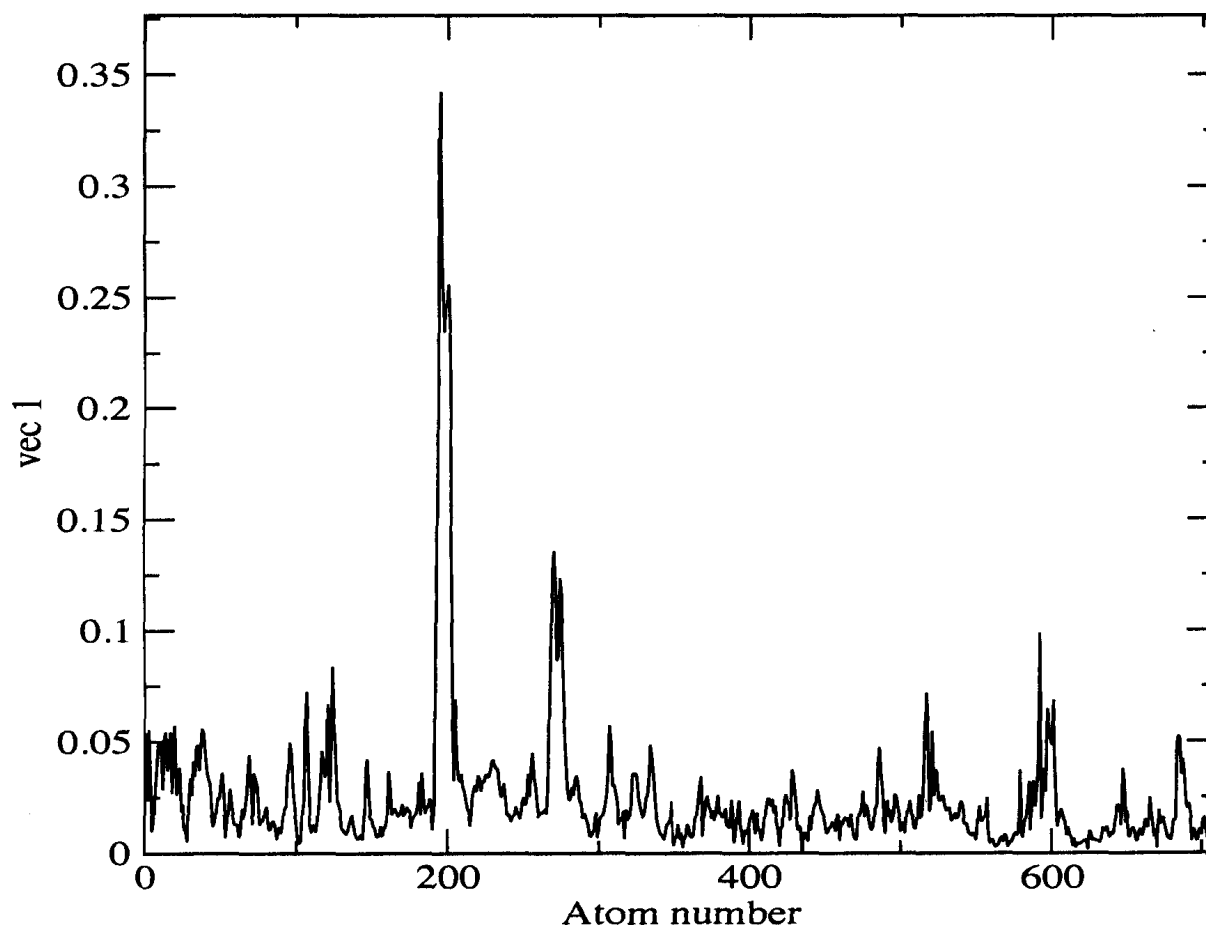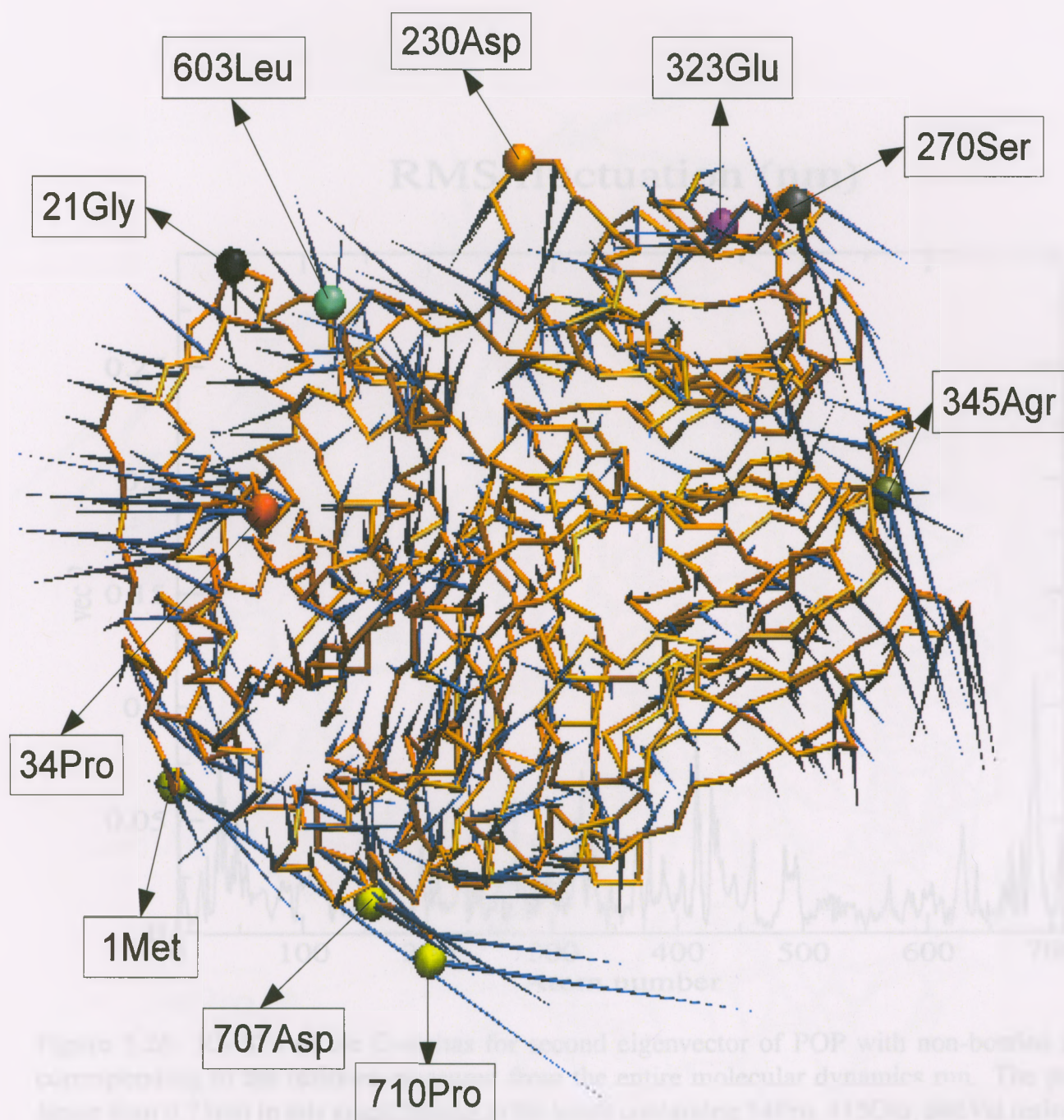
Figure 5.25: RMSFs of the C-alpha atoms for first eigenvector of POP with free ZPP corresponding to the residues measured from the entire molecular dynamics run. The residues mentioned in Fig. 5.24 are labeled here. The arrows show direction and magnitude of movements. Note that the ZPP atom is not shown for clarity.

RMS fluctuation (nm)
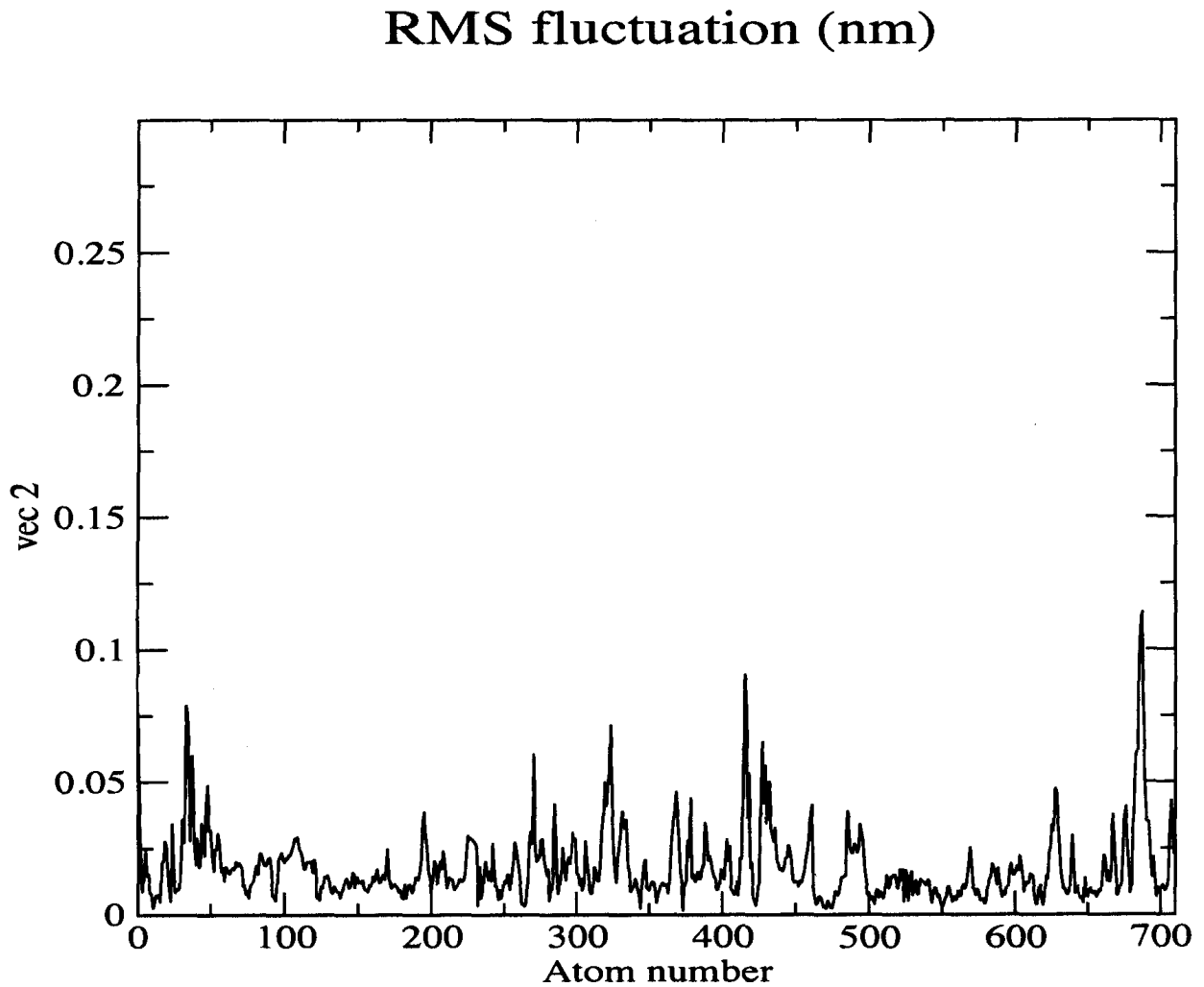
Figure 5.26: RMSFs of the C-alphas for second eigenvector of POP with non-bonded ZPP corresponding to the residues measured from the entire molecular dynamics run. The peaks larger than 0.75nm in this graph belong to the loops containing 34Pro, 415Glu, 686Val residues. These residues are labeled in Fig. 5.27. Note that the ZPP atom is not shown for clarity.
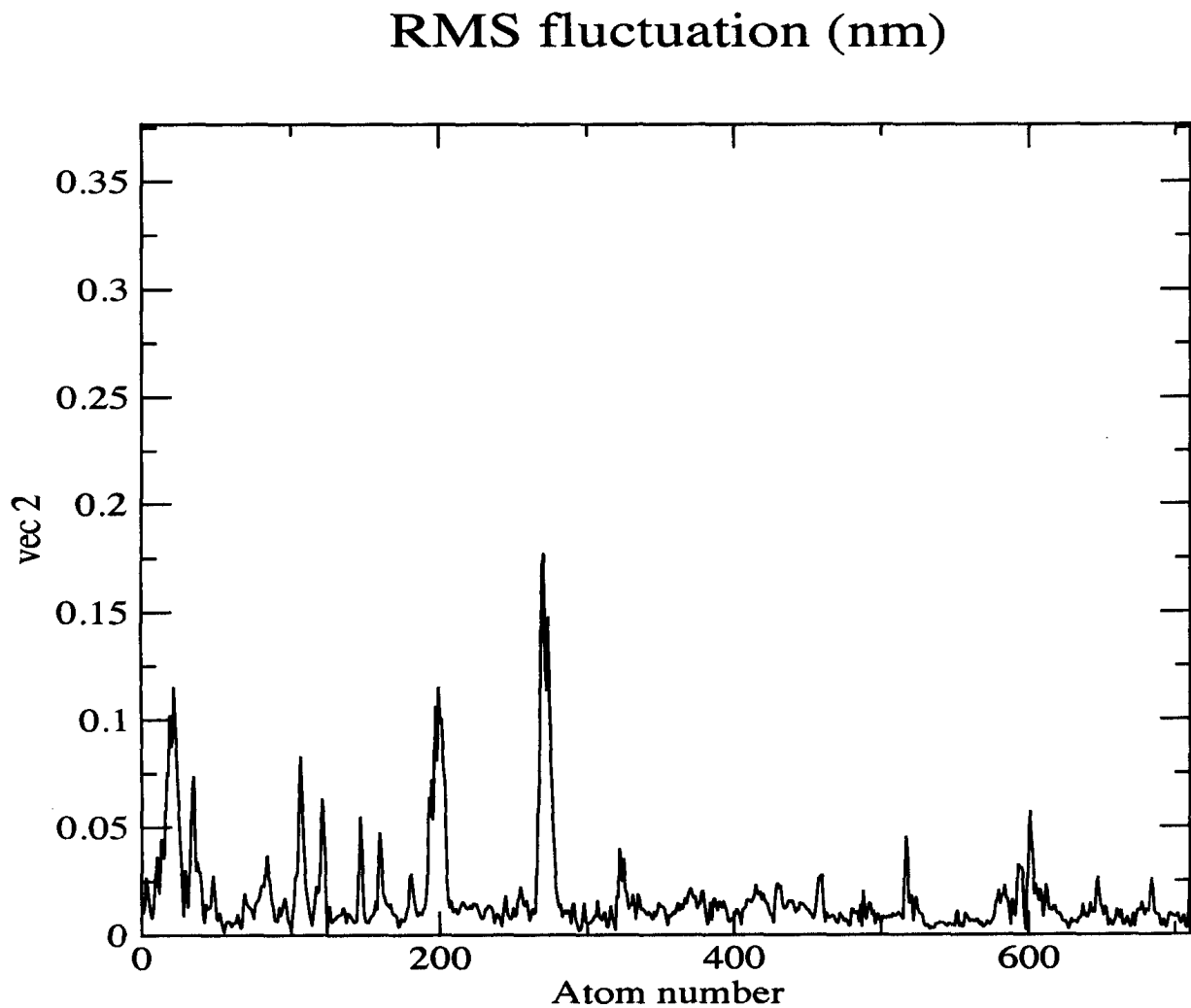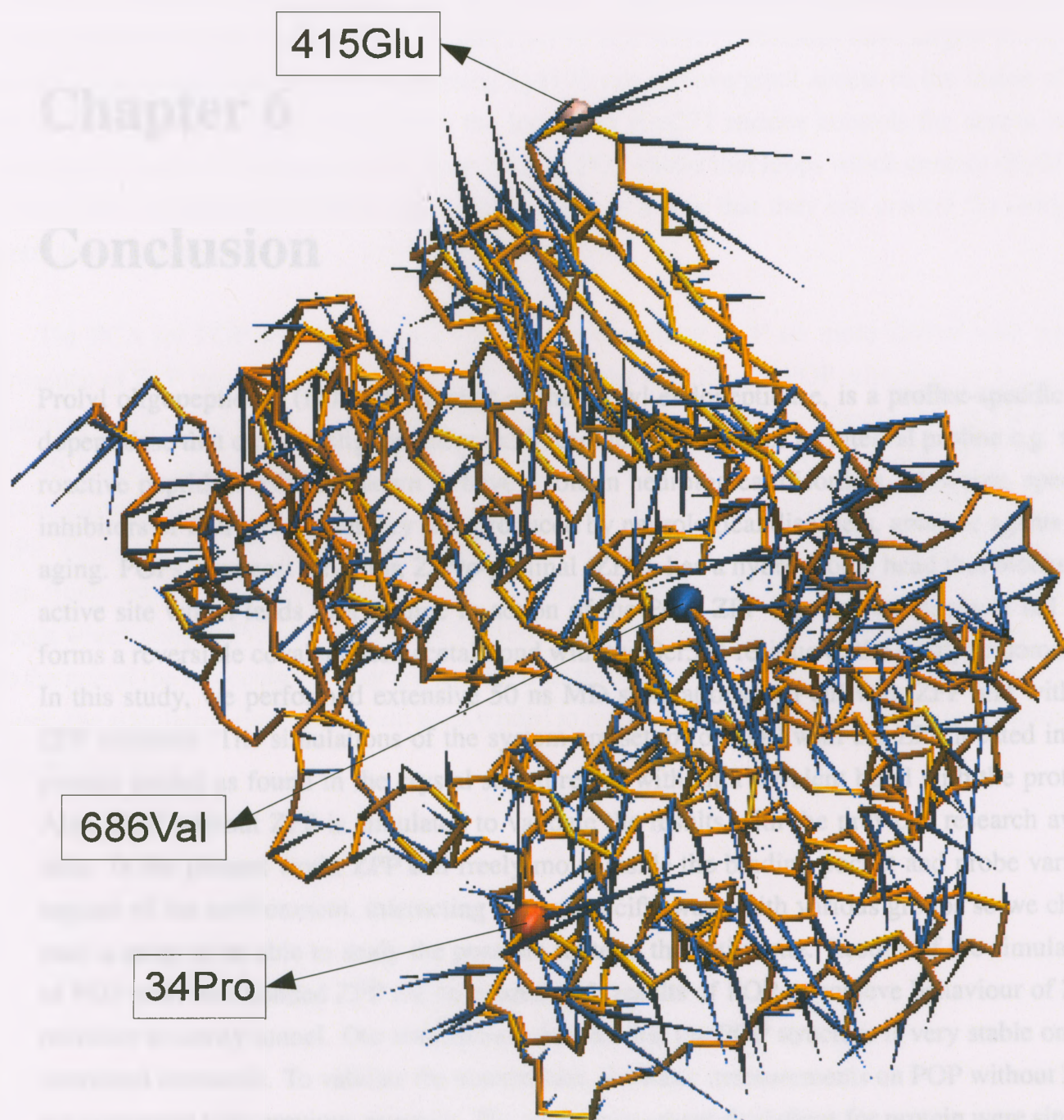
Figure 5.27: RMSFs of the C-alphas for second eigenvector of POP with free ZPP correspond-ing to the residues measured from the entire molecular dynamics run. The residues mentioned in Fig. 5.26 are labeled here. Note that ZPP is not shown for clarity of the figure.

Chapter 6

Conclusion

Prolyl oligopeptidase (POP), a member of the prolyl endopeptidase, is a proline-specific endopeptidase that cleaves oligopeptides (<30 residues) at C-side of an internal proline e.g. neuroactive peptides. POP is known to have a role in neurological disorders. However, specific inhibitors of POP revert memory loss produced by neurological disorders, amnesic agents and aging. POPs strongest inhibitor, Z-Pro-prolinal (ZPP), has a hydrophobic head that blocks the active site which leads to a change in action of the POP. ZPP also has an aldehyde tail that forms a reversible covalent hemiacetal bond with the Ser554 residue of the catalytic domain.

In this study, we performed extensive 50 ns MD simulations of POP with ZPP and without ZPP inhibitor. The simulations of the system are setup for POP with the ZPP located in the protein pocket as found in the crystal structure but without a covalent bond with the protein. Also, POP without ZPP is simulated to validate the results with the previous research available. In the present work, ZPP can freely move inside the binding pocket and probe various regions of the environment, interacting in non-specific ways with various groups so we chose such a setup to be able to study the possible route to the active site. Results of the simulation of POP with non-bonded ZPP are compared with results of POP to achieve behaviour of ZPP inhibitor in cavity tunnel. Our simulations showed that the POP structure is very stable on the simulated timescale. To validate the simulations, the basic measurements on POP without ZPP are compared with previous research. The root mean square deviations for protein were similar to those obtained by [95] Kaszuba et al. A slight difference was observed due to the difference in temperature of the systems simulated in the presented work (303 Kelvin) and [95] Kaszuba et al. (310 kelvin). The radius of gyration for C-alpha carbons were compared and found to be similar while showing the POP was stable in whole simulation. More basic characteristic for POP with ZPP were compared with POP without ZPP. These results demonstrate that the protein is stable in whole simulation period when the ZPP is added to system. A principal component analysis was performed for POP and POP with ZPP. Although simulations on POP with

ZPP have previously been done by K. Kaszuba *et al.*, but the PCA for POP has not been yet accomplished. Here, the results from PCA are analyzed in order to get a better understanding of the movements of the molecules. The first mode of motions obtained from PCA analysis for POP indicates that the loops which include Lys196 and Asn271 residues have largest movements. The motions of the loop containing Lys196 can remove/grant access to the inside of the cavity through the velcro-rip, while the loop with Asn271 residue controls the access to β-propeller tunnel. The second mode from PCA of POP shows that loops which contain Gly21 and Lys196 residues move towards/away each other in a way that they can control the entry point to the velcro-rip.

The PCA for POP with ZPP shows that the movements of POP are more limited with the presence of ZPP inhibitor which can disable most of the activities of POP related to its unique pathways. The first mode of motions from PCA for POP with ZPP indicates that the loops which include residues Gly21, Pro34, 230Asp, 270Ser, 323Glu, 345Agr and 603Leu have some movements but these motions are relatively smaller when compared with POP without ZPP. These movements do not seem large enough to open either of the pathways so that ZPP can freely escape the cavity. The second mode of motions for POP with ZPP show that the loops which contain Pro34, Glu415 and Val686 residues have the largest movements compared to the rest of the molecule but they are smaller than those related to POP without ZPP. As future work, the simulation run time should be extended to longer periods to get a better understanding of the smulation.

Bibliography

[1] Gromacs. Gromacs Fast Flexible Free, 2011.

[2] B. Aaberts, D. Bray, A. Johnson, and J. Lewis. *Essential Cell Biology*. Garland Publishing, New York and London, 1998.

[3] R. Abseher, L. Horstink, C. W. Hilbers, and M. Nilges. Essential spaces defined by nmr structure ensembles and molecular dynamics simulation show significant overlap. *Proteins: Struct. Funct. Genet.*, 31:370–382, 1998.

[4] B. Albert, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology of the Cell*. Garland Publishing, New York, 3rd edition, 1994.

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

[6] A. Amadei, M. A. Ceruso, and A. D. Nola. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Struct. Funct. Genet.*, 36:419–424, 1999.

[7] A. Amadei, B. L. de Groot, M. A. Ceruso, M. Paci, A. D. 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.

[8] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins: Struct. Funct. Genet*, 17:412–425, 1993.

[9] A. Amadei, A. B. M. Linssen, B. L. de Groot, D. M. F. van Aalten, and H. J. C. Berendsen. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.*, 13:615–626, 1996.

[10] P. Anderson. Amino acid: Categories and function. school work helper, November 2010.

[11] P. Auffinger, T.E. Cheatham, and A.C. Vaiana. Spontaneous formation of kcl aggregates in biomolecular simulations: A force field issue? *J. Chem. Theory Comput.*, 3:1851–1859, 2007.

[12] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14:5355–5373, 1975.

[13] G. S. Ayton, W. G. Noid, and G. A. Voth. Multiscale modeling of biomolecular systems: In serial and in parallel. *Curr. Opin. Struct. Biol.*, 17:192–198, 2007.

[14] S. C. Baker, N. F. W. Saunders, A. C. Willis, S. J. Ferguson, J. Hajdu, and V. Fülöp. Cytochrom cd1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to beta-propeller folds. *J. Mol. Biol.*, 269:440–455, 1997.

[15] K. K. Baldridge and C. Amoreira. Molecular mechanics. AccessScience, ©McGraw-Hill Companies, 2008.

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

[17] R. Baron, D. Trzesniak, A. H. de Vries, A. Elsener, S. J. Marrink, and W. F. van Gunsteren. Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. *Chem. Phys. Chem.*, (8):452, 2007.

[18] C. P. Barrett, B. A. Hall, and M. E. M. Nobel. Dynamite: A simple way to gain insight into protein motions. *Acta Cryst.*, 60(1):2280–2287, 2004.

[19] M. Bartlam, G. Wang, H. Yang, R. Gao, X. Zhao, G. Xie, Sh. Cao, Y. Feng, and Z. Rao. Crystal structure of an acylpeptide hydrolase/esterase from aeropyrum pernix k1. *Structure*, 12:1481–1488, 2004.

[20] M. Bartlam, G. Wang, H. Yang, X. Zhao, G. Xie, Sh. Cao, Y. Feng, and Z. Rao. Crystal structure of an acylpeptide hydrolase/esterase from pernix k1. *Structure*, 12:1481–1488, 2004.

[21] C.I. Bayly, P. Cieplak, W.D. Cornell, and Kollman P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model. *J. Phys. Chem.*, 97:10269–10280, 1993.

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

[23] B. Brooks and M. Karplus. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, 80:6571–6575, 1983.

[24] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity-rescaling. *J. Chem. Phys*, 126, 2007.

[25] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev*, (55):2471–2474, 1985.

[26] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.

[27] D. A. Case, T. E. Cheatman, T. Darden, H. Gohlke, R. Luo, Jr. Merz, K. M., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J. Comp. Chem.*, 26:1668–1688, 2005.

[28] J. J. Cazzulo. *Chagas disease. In: Proteases of Infectious Agents (Dunn. B. M. Ed.).* Academic Press, London, 1999.

[29] S. Chevalier, P. Goeltz, P. Thibault, D. Banville, and J. Gagnon. Characterization of prolyl oligopeptidase from flavobacterium meningosepticum. *J. Biol. Chem.*, 267:8192–8199, 1992.

[30] M. Christen, P. H. Hunenberger, D. Bakowies, R. Baron, Burgi R., D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren. The gromos software for biomolecular simulation: Gromos05. *J. Comp. Chem.*, 26:1719–1751, 2005.

[31] J. B. Clarage, T. Romo, B. K. Andrews, B. M. Pettitt, and G. N. P. Jr. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA*, 92:3288–3292, 1995.

[32] D. COTTRELL1 and P.F. TUPPER. Energy drift in molecular dynamics simulations. *BIT Numerical Mathematics*, 47:507–523, may 2007.

[33] D. F. Cunningham and B OConnor. Proline specific peptidases. *Biochem. Biophys. Acta*, 1343:160–186, 1997.

[34] M. Cygler, J. D. Schrag, J. L. Sussman, M. Harel, I. Silman, and M. K. et al. Gentry. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. *Protein Sci.*, 2:366 – 382, 1993.

[35] T. Darden, D. York, and L. G. Pedersen. Particle mesh ewald: An n*log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[36] F. David, A. M. Bernard, M. Pierres, and D. Marguet. Identification of serine 624, aspartic acid 702 and histidine 734 as the catalytic triad residues of mouse dipeptidyl-pepti- dase iv (cd26). a member of a novel family of nonclassical serinehydrolases. *J. Biol. Chem.*, 268:17247–17252, 1993.

[37] B. L. de Groot, A. Amadei, R. M. Scheek, N. A. J. van Nuland, and H. J. C. Berendsen. An extended sampling of the configurational space of hpr from e. coli. *Proteins: Struct. Funct. Genet.*, 26:314–322, 1996.

[38] 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. Biomol. Struct. Dyn.*, 13:741–751, 1996.

[39] B. L. de Groot, D. M. F. van Aalten, A. Amadei, and H. J. C Berendsen. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys. J.*, 71:1707–1713, 1996.

[40] 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.*, 286:1241–1249, 1999.

[41] S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. i. lattice sums and dielectric constants. *Proc. Roy. Soc. Lond. A*, 373:27–56, 1980.

[42] S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulations of electrostatic systems in periodic boundary conditions. ii. equivalence of boundary conditions. *Proc. Roy. Soc. Lond. A*, 373:57–66, 1980.

[43] B. L. deGroot, S. Hayward, D. M. F. v. Aalten, A. Amadei, and H. J. C. Berendsen. Domain motions in bacteriophage t4 lysozyme; a comparison between molecular dynamics and crystallographic data. *Proteins: Struct. Funct. Genet.*, 31:116–127, 1998.

[44] T. Diefenthal, H. Dargatz, V. Witte, G. Reipen, and I. Svendsen. Cloning of proline-specific endopeptidase gene from flavobacterium meningosepticum: expression in escherichia coli and purification of the heterologous protein. *Appl. Microbiol. Biotechnol.*, 40:90–97, 1993.

[45] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory*. Springer, 1990.

[46] D. J. Drucker. Minireview: the glucagon-like peptides. *Endocrinology*, 142:521–527, 2001.

[47] R. Elber and M. Karplus. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science*, 235:318–321, 1989.

[48] F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *Europhys. Lett.*, 26(8):583–588, 1994.

[49] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh emald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[50] P. P. Ewald. The calculation of optical and electrostatic grid potential. *Ann. Phys. (Leipzig)*, 64:253, 1921.

[51] H. R. Faber, C. R. Groom, H. M. Baker, W. T. Morgan, A. Smith, and E. N. Baker. 1.8 å structure of the c-terminal domain of rabbit serum hemopexin. *Structure*, 3:551–559, 1995.

[52] T. C. Farries, A. Harris, A. D. Auffret, and A. Aitken. Removal of n-acetyl groups from blocked peptides with acylpeptide hydrolase. stabilization of the enzyme and its application to protein sequencing. *Eur. J. Biochem.*, 196:679–685, 1991.

[53] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1 of *Computational Science Series*. ACADEMIC PRESS, New York, second edition, 2002.

[54] V. Fulop. Prolyl oligopeptidase from porcine muscle with covalently bond inhibitor z-pro-prolinal. Protein Data Bank (PDB), 2009.

[55] V. Fülöp, Z. Böcskei, and L. Polgár. Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell*, 94:161–170, 1998.

[56] V. Fülöp, Z. Szeltner, and L. Polgár. Catalysis of serine oligopeptidases is controlled by a gating filter mechanism. *EMBO J.*, 1:277–281, 2000.

[57] W. Gade and J. L. Brown. Purification and partial characterization of alpha-n-acylpeptide hydrolase from bovine liver. *J. Biol. Chem.*, 253:5012–5018, 1978.

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

[59] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer, New York, 3rd edition, 2003.

[60] N. Go. A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. *Biophysical Chemistry*, 35:105–112, 1990.

[61] N. Go, T. Noguti, and T. Nishikawa. Dynamicsofasmallglobularproteinintermsof low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*, 80:3696–3700, 1983.

[62] H Grubmu ller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E.*, 52:2893–2906, 1995.

[63] P. Gumbsch. *Multiscale Moldelling of Plasticity and Fracture by Means of*, volume 522 of *CISM Courses and lectures*. International Centre for Mechanical Science, 1 edition, 2010.

[64] M. N. Harris, J. D. Madura, L. J. Ming, and V. J. Harwood. Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile pyrococcus furiosus. *J. Biol. Chem.*, 279:19310–19317, 2001.

[65] V. J. Harwood and H. J. Schreier. Prolyl oligopeptidase from pyrococcus furiosus. *Methods Enzymol.*, 330:445–454, 2001.

[66] S. Hayward. Normal mode analysis of biological molecules. Computational Biochemistry and Biophysics (Marcel Dekker Inc, New York.), 2001.

[67] S. Hayward. *Normal mode analysis of biological molecules. In Computational Biochemistry and Biophysics*. Marcel Dekker Inc, New York, 2001.

[68] S. Hayward and B. L. de Groot. *Molecular Modelling of Protein*, volume 443 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ, 2008.

[69] S. Hayward, A. Kitao, and N. Go. Harmonicity and anharmonicity in protein dynamics: a normal modes and principal component analysis. *Proteins: Struct. Funct. Genet.*, 23:177–186, 1995.

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

[71] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[72] B. et al. Hess. *GROMACS Groningen Machine for Chemical Simulations*. The GROMACS development teams at the Royal Institute of Technology and Uppsala University, 4.5.4 edition.

[73] D. M. Heyes. Molecular dynamics at constant pressure and temperature. *J. Chem. Phys.*, 82:285–301, 1983.

[74] H. Hiramatsu, K. Kyono, Y. Higashiyama, C. Fukushima, H. Shima, S. Sugiyama, K. Inaka, A. Yamamoto, and R. Shimizu. The structure and function of human dipeptidyl peptidase iv, possessing a unique eight-bladed beta-propeller fold. *Biochem. Biophys. Res. Commun.*, 302:849–854, 2003.

[75] R.W. Hockney and J.W. Eastwood. *Computer Simulations Using Particles*. McGraw-Hill, New York, 1981.

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

[77] V. K. Hopsu-Havu and G. G. Glenner. A new dipeptide naphthylamidase hydrolyzing glycyl-prolyl-beta-naphthlamide. *Histochemie*, 7:197–201, 1966.

[78] W Humphrey, A Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[79] P.H. Hunenberger. Thermostat algorithms for molecular dynamics simulations. *Adv. Polymer. Sci.*, 173:105–149, 2005.

[80] Y. Ikehara, S. Ogata, and Y. Misumi. Dipeptidyl-peptidase iv from rat liver. *Methods Enzymol.*, 244:215–227, 1994.

[81] K. Ishikawa, H. Ishida, Y. Koyama, Y. Kawarabayasi, J. Kawahara, and E. et al. Matsui. Acylamino acid-releasing enzyme from the thermophylic archaeon pyrococcus horikoshii. *J. Biol. Chem.*, 273:17726–17731, 1998.

[82] T. Ishino, S. Ohtsuki, K. Homma, and S. Natori. cdna cloning of mouse prolyl endopeptidase and its involvement in dna synthesis by swiss 3t3 cells. *J. Biochem.*, 123:540–545, 1998.

[83] R. Ishitani, M. Tanaka, K. Sunaga, and D. M. Chang. Nuclear localization of overexpressed glyceraldehyde-3-phosphate dehydrogenase in cultured cerebellar neurons undergoing apoptosis. *Mol. Pharmacol.*, 53:701–707, 1998.

[84] Y Ji. Molecular biology of the cell. Advanced User Training Database Advanced User Training Database Advanced user Training Database, December 2009.

[85] R. C. Johnson, D. Zhu, H. G. AugustiVoss, and B. U. Pauli. Lung endothelial dipeptidyl-peptidase iv is an adhe- sion molecule for lungmetastatic rat breast and prostate carcinomacells. *J. Cell Biol.*, 121:1423–1432, 1993.

[86] W. M. Jones, L. R. Manning, and J. M. Manning. Enzymic cleavage of the blocked amino terminal residues of peptides. *Biophys. Res. Commun.*, 139:244–250, 1986.

[87] W. M. Jones, A. Scaloni, F. Bossa, A. M. Popowicz, O. Schneewind, and J. M. Manning. Genetic relationship between acylpeptide hydrolase and acylase, two hydrolytic enzymes with similar binding but different catalytic specificities. *Proc. Natl. Acad. Sci. USA*, 88:2194–2198, 1991.

[88] W. M. Jones, A. Scaloni, and J. M. Manning. Acylaminoacyl peptidase. *Methods Enzymol*, 244:227–231, 1994.

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

[90] W. L. Jorgensen, S. D. 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.

[91] 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.

[92] Ju. Tetrapeptide structural formulae. wikipedia, 11 2010.

[93] A. Kanatani, T. Masuda, T. Shimoda, F. Misoka, X. S. Lin, and T. et al. Yoshimoto. Protease ii from escherichia coli: sequencing and expression of the enzyme gene and characterization of the expressed enzyme. *J. Biochem.*, 110:315–320, 1991.

[94] A. Kanatani, T. Yoshimoto, A. Kitazono, T. Kokubo, and T. Tsuru. Prolyl endopeptidase from aeromonas hydrophila: cloning, sequencing, and expression of the enzyme gene, and characterization of the expressed enzyme. *J. Biochem.*, 113:790–796, 1993.

[95] K. Kaszuba, T. Rog, JF. St Pierre, PT. Mannisto, M. Karttunen, and A. Bunker. Molecular dynamics study of prolyl oligopeptidase with inhibitor in binding cavity. *Taylor and Francis*, 20(7-8):595–609, 2009.

[96] A. Kimura, I. Yoshida, N. Takagi, and T. Takahashi. Structure and localization of the mouse prolyl oligopeptidase gene. *J. Biol. Chem.*, 274:24047–24053, 1999.

[97] A. Kitao and N. Go. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.*, 9:143–281, 1999.

[98] A. Kitao, S. Hayward, and N. Go. Energy landscape of a native protein: jumping-among-minima model. *Proteins: Struct. Funct. Genet.*, 33:496–517, 1998.

[99] A. Kitao, F. Hirata, and N. Go. 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. *J. Chem. Phys.*, 158:447–472, 1991.

[100] W. Kohn and L. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 4(140):1133–1138, 1965.

[101] M. Koida and R. Walter. Post-proline cleaving enzyme. purification of this endopeptidase by affinity chromatography. *J. Biol. Chem.*, 251(7593-7599), 1976.

[102] A. Laio, J. Vande, J. Vondele, and U. Rothlisberger. A hamiltonian electrostatic coupling scheme for hybrid car-parrinello simulations. *J. Chem. Phys*, 116:6941–6948, 2002.

[103] H. Lee and H. Cai. Ewald summation for coulomb interactions in a periodic supercell. Technical report, Department of Mechanical Engineering, Stanford University, 2009.

[104] O. Lenz. Lennard-jonse potential. Wikipedia, March 2007.

[105] M. Levitt. *Energy Calculations and Dynamics Program; 1990.* Molecular Applications Group: Stanford, 1990.

[106] M. Levitt and P. S. Sander, C. aznd Stern. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int. J. Quant. Chem.*, 10:181–199, 1983.

[107] J. Li, P. Brick, M. C. Ohare, T. Skarzynski, L. F. Lloyd, and V. A. et al. Curry. Structure of full-length porcine synovial collagenase reveals a c-terminal domain containing a calcium-linked, four-bladed beta-propeller. *Structure*, 3:341–349, 1995.

[108] C. N. Likos. Effective interactions in soft condensed matter physics. *Physics Reports*, (348):267–439, 2001.

[109] A. Lyubartsev, A. Mirzoev, L. Chen, and A. Laaksonen. Systematic coarse-graining of molecular models by the newton inversion method. *Faraday Discuss.*, 144:43–56, 2010.

[110] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: a reverse monte carlo approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 52:3730–3737, 1995.

[111] M. Maes, F. Goossens, S. Scharpé, H. Y. Meltzer, P. D'Hondt, and P. Cosyns. Lower serum prolyl endopeptidase enzyme activity in major depression: further evidence that peptidases play a role in the pathophysiology of depression. *Biol. Psychiatry*, 35:545–552, 1994.

[112] Y Marabet. Peptide bond formation. Wikipedia, August 2007.

[113] O. Marques and Y. H. Sanejouand. Hinge-bendingmotionincitratesynthasearising from normal mode calculations. *Proteins*, 23:557–560, 1995.

[114] D. Marx and J. Hutter. Ab initio molecular dynamics: Theory and implementation. *Modern Methods and Algorithms of Quantum Chemistry*, 1:301–449, 2000.

[115] K. Mcluskey, N. G. Paterson, N. D. Bland, N. W. Isaacs, and J. C. Mottram. Crystal structure of leishmania major oligopeptidase b gives insight into the enzymatic properties of a trypanosomatid virulence factor. *J. Biol. Chem.*, 285:39249, 2010.

[116] F. J. Medrano, J. Alonso, J. L. García, A. Romero, W. Bode, and F. X. Gomis-Rüth. Stucture of proline iminopeptidase from xanthomonas campestris pv. citri: a prototype for the prolyl oligopeptidase family. *EMBO J.*, 17:1–9, 1998.

[117] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink. The martini coarse grained force field: extension to proteins. *J. Chem. Theory Comput.*, (4):819, 2008.

[118] Y. Mrabet. Structure of amino acids. Wikipedia, August 2007.

[119] F. Muller-Plathe. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem*, 3:754–769, 2002.

[120] T. Murtola. Coarse-grained model for phospholipid/cholesterol bilayer. Master's thesis, Helsinki University of Technology, 2004.

[121] P. Niemela. Computational modeling of lipid bilayers with sphingomyelin and sterols. Academic Dissertation, 2007.

[122] W. Noid, J. Chu, G. Ayton, and G. Voth. Multiscale coarse-graining and structural correlations: connections to liquid-state theory. *J. Phys. Chem. B*, 111(16):4116–4127, 2007.

[123] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.

[124] S. Ogata, Y. Misumi, and Y. Ikehara. Primary structure of rat liver dipeptidyl-peptidase iv deduced from its cdna and identification of the nh2-terminal signal sequence as the membrane-anchoring domain. *J. Biol. Chem.*, 264:3596–3601, 1989.

[125] S. Ohtsuki, K. Homma, S. Kurata, and S. Natori. Molecular cloning of cdna for sarcophaga prolyl endopeptidase and characterization of the recombinant enzyme produced by an e. coli expression system. *Insect. Biochem. Mol. Biol.*, 27:337–343, 1997.

[126] D. I. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, and S. M. et al. Franken. The a/b hydrolase fold. *Protein Eng*, 5:197–211, 1992.

[127] M. Pacaud. Purificatioön of protease ii from escherichia coli by affinity chromatography and separation of two enzyme species from cells harvested at late log phase. *Eur. J. Biochem.*, 64:99–104, 1976.

[128] M. Pacaud and C. Richaud. Protease ii from escherichia coli. *J. Biol. Chem.*, 250:7771–7779, 1975.

[129] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*, volume 16 of *International Series of Monographs On Chemistry*. Oxford University Press, 1989.

[130] M. Patra and M. Karttunen. Systematic comparison of force fields for microscopic simulations of nacl in aqueous solutions: Diffusion, free energy of hydration, and structural properties. *J. Comp. Chem.*, 25:678–689, 2004.

[131] A. Petit, H. Barelli, P. Morain, and F. Checler. Novel pro- line endopeptidase inhibitors do not modify abeta40/42 for- mation and degradation by human cells expressing wild-type and swedish mutated beta-amyloid precursor protein. *Biochem. Biophys. Res. Commun.*, 257:657–661, 2000.

[132] L. Polgar. The prolyl oligopeptidase family. *CMLS, Cellular and Molecular Life Sciences*, 59:349–362, 2002.

[133] B. Qian, A. R. Ortiz, and D. Baker. Improvement of comparative model accuracy by free-energy optimization along principal components of natural structural variation. *Proc. Natl. Acad. Sci. USA*, 101:15346–15351, 2004.

[134] G. Radhakrishna and F. Wold. Purification and characterization of alpha-n-acylpeptide hydrolase from rabbit muscle. *J. Biol. Chem.*, 264:11076–11081, 1989.

[135] V. Raphel, T. Giardina, L. Guevel, J. Perrier, L. Dupuis, and X. J. et al. Gou. Cloning, sequencing and further characterization of acylpeptide hydrolase from porcine intestinal mucosa. *Biochem. Biophys. Acta*, 1432:371–381, 1999.

[136] N. D. Rawling, L. Polgar, and Barrett. A. J. A new familyofserine-typepeptidases relatedto prolyloligopeptidase. *Biochem. J. Lett.*, 279:907–911, 1991.

[137] N. D. Rawlings, Polgár L., and Barrett A. J. A new family of serine-type peptidases related to prolyl oligopeptidase. *Biochem. J.*, 279:907–908, 1991.

[138] D.K. Remler and P.A. Madden. Molecular dynamics without effective potentials via the car-parrinello approach. *Mol. Phys.*, (70):921–966, 1990.

[139] D. Rennex, B. A. Hemmings, J. Hofsteenge, and S. R. Stone. cdna cloning of porcine brain prolyl endopeptidase and identification of the active-site seryl residue. *Biochemistry*, 30:2195–2203, 1991.

[140] K. A. Robinson, D. A. Bartley, F. T. Robb, and H. J. Schreier. A gene from the hyperthermophile pyrococcus furiosus whose deduced product is homologous to members of the prolyl oligopeptidase family of proteases. *Gene*, 152:103–106, 1995.

[141] V. Ruehle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput.*, 5:3211–3223, 2009.

[142] G. Santangelo, A. Di Matteo, F. Muller-Plathe, and G. Milano. From mesoscale back to atomistic models: A fast reverse-mapping procedure for vinyl polymer chains. *J. Phys. Chem. B*, 111:2765, 2007.

[143] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, and J.A. Montgomery Jr. The general atomic and molecular electronic structure system. *J. Comp. Chem.*, 14:1347–1362, 1993.

[144] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B*, 17(3):1302–1322, 1978.

[145] J. D. Schrag and M. Cygler. Lipases and alpha/beta hydrolase fold. *Methods Enzymol*, 284:85–107, 1997.

[146] M. Shinoda, K. Toide, I. Ohsawa, and S. Kohsaka. 41shinoda m., toide k., ohsawa i. and kohsaka s. (1997) specific inhibitor for prolyl endopeptidase suppresses the generation of amiloid b protein in ng108-15 cells. *Biochem. Biophys. Res. Commun*, 235:641–645, 1997.

[147] Y. Shirasawa, T. Osawa, and A. Hirashima. Molecular cloning and characterization of prolyl endopeptidase from human t cells. *J. Biochem.*, 115:724–729, 1994.

[148] G. D. Smith, E. Ciszak, L. A. Magrum, W. A. Pangborn, and R. H. Blessing. R6 hexameric insulin complexed with m-cresol or resorcinol. *Biochem. Biophys. Res. Commun.*, 56:1541–1548, 2000.

[149] L. C. Smith, F. Faustinella, and L. Chan. Lipases: three- dimensional structure and mechanism of action. curr. opin. *Struct. Biol.*, 2:490–496, 1992.

[150] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: Fast, flexible, and free. *J. Comp. Chem.*, 26(16):1701–1718, 2005.

[151] M. Srikant. Potential due to point charges, September 2000.

[152] S. R. Stone, D. Rennex, P. Wikstrom, E. Shaw, and J. Hofsteenge. Inactivation of prolyl endopeptidase by a peptidylchloromethane. kinetics of inactivation and identification of sites of modification. *Biochem. J.*, 276:837–840, 1991.

[153] W. C. Swope, H. C. Andersen, 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.

[154] C. Tiruppathi, V. Ganapathy, and F. H. Leibach. Evidence for tripeptide-proton symport in renal brush border membrane vesicles. studies in a novel rat strain with a genetic absence of dipeptidyl peptidase iv. *J. Biol. Chem.*, 265:2048–2053, 1990.

[155] Fülöp V. and Jones D. T. beta-propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.*, 9:715–721, 1999.

[156] D. M. F. van Aalten, B. L. de Groot, H. J. C. Berendsen, J. B. C. Findlay, and A. Amadei. A comparison of techniques for calculating protein essential dynamics. *J. Comp. Chem.*, 18:169–181, 1997.

[157] D. M. F. van Aalten, B. L. de Groot, H. J. C. Berendsen, J. B. C. Findlay, and A. Amadei. A comparison of techniques for calculating protein essential dynamics. *J. Comp. Chem.*, 18:169–181, 1997.

[158] G. Vanhoff, F. Goossens, L. Hendriks, I. De Meester, D. Hendriks, and G. et al. Vriend. Cloning and sequence analysis of the gene encoding human lymphocyte prolyl endopeptidase. *Gene (Amst)*, 149:363–366, 1994.

[159] I. Vattulainen and M Karttunen. *Handbook of Theoretical and Computational Nanotechnology*, chapter 29, pages 1–57. American Scientific Publishers, 2005.

[160] A. Villa, N. F. A. van der Vegt, and C. Peter. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Phys. Chem. Chem. Phys.*, 11:2068, 2009.

[161] T. Vilsboll, T. Krarup, C. F. Deacon, S. Madsbad, and J. J. Holst. Reduced postprandial concentrations of intact biologically active glucagon-like peptide 1 in type 2 diabetic patients. *Diabetes*, 50:609–613, 2001.

[162] J. V. Vondele. *Extending length and time scales of ab initio molecular dynamics simulations*. PhD thesis, Engineer in Physics, Ghent University, August 2001.

[163] R. Walter, H. Shlank, J. D. Glass, I. L. Schwartz, and T. D. Kerenyi. Leucylglycinamide released from oxytocin by humanuterineenzyme. *Science*, 173:827–829, 1971.

[164] W. R. Welches, K. B. Brosnihan, and C. M. Ferrario. A comparison of the properties and enzymatic activities of three angiotensin processing enzymes: angiotensin converting enzyme, prolyl endopeptidase and neutral endopeptidase. *Life Sci.*, 52:1461–1480, 1993.

[165] M.J Wu. Protein ligand binding. Wikipedia, December 2008.

[166] T. Yoshimoto, A. Kanatani, T. Shimoda, T. Inaoka, T. Kobubo, and D. Tsuru. Prolyl endopeptidase from flavobac- terium meningosepticum: cloning and sequencing of the enzyme gene. *J. Biochem.*, 110:873–878, 1991.

[167] T. Yoshimoto, K. Miyazaki, N. Haraguchi, A. Kitazono, T. Kabashima, and K. Ito. Cloning and expression of the cdna encoding prolyl oligopeptidase (prolyl endopeptidase) from bovine brain. *Biol. Pharm. Bull.*, 20:1047–1050, 1997.

[168] Y. Yoshimoto, K. Kado, F. Matsubara, N. Koriyama, H. Kaneto, and D. Tsura. Specific inhibitors for prolyl endopeptidase and their anti-amnesic effect. *J. Pharmacobiodyn.*, 10:730–735, 1987.

Appendix A

License 1

Appendix B

License 2

This is a License Agreement between sepideh soltani ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	sepideh soltani
Customer address	1252 baird st london, ON n6h0g5
License number	2811370299053
License date	Dec 17, 2011
Licensed content publisher	Elsevier
Licensed content publication	Cell
Licensed content title	Prolyl Oligopeptidase: An Unusual β-Propeller Domain Regulates Proteolysis
Licensed content author	Vilmos Fülöp,Zsolt Böcskei,László Polgár
Licensed content date	24 July 1998
Licensed content volume number	94
Licensed content issue number	2
Number of pages	10
Start Page	161
End Page	170
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	print
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	MOLECULAR DYNAMICS ANALYSIS OF THE BEHAVIOUR OF PROLYL OLIGOPEPTIDASE (POP) IN THE PRESENCE OF Z-PRO-PROLINAL INHIBITOR
Expected completion date	Dec 2011
Estimated size (number of pages)	114
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD
Terms and Conditions	

Appendix C

License 3

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 21, 2011

This is a License Agreement between sepideh soltani ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	sepideh soltani
Customer address	1252 baird st
	london, ON n6h0g5
License number	2810851392290
License date	Dec 16, 2011
Licensed content publisher	Elsevier
Licensed content publication	Structure
Licensed content title	Crystal Structure of an Acylpeptide Hydrolase/Esterase from*Aeropyrum pernix*K1
Licensed content author	Mark Bartlam,Ganggang Wang,Haitao Yang,Renjun Gao,Xiaodong Zhao,Guiqiu Xie,Shuigui Cao,Yan Feng,Zihe Rao
Licensed content date	August 2004
Licensed content volume number	12
Licensed content issue number	8
Number of pages	8
Start Page	1481
End Page	1488
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	print
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	MOLECULAR DYNAMICS ANALYSIS OF THE BEHAVIOUR OF PROLYL OLIGOPEPTIDASE (POP) IN THE PRESENCE OF Z-PRO-PROLINAL INHIBITOR
Expected completion date	Dec 2011
Estimated size (number of pages)	114
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD
Terms and Conditions	

104

Appendix D

License 4

Copyright Clearance Center RightsLink®

Home | Account Info | Help

Taylor & Francis
Taylor & Francis Group

Title: Molecular dynamics study of prolyl oligopeptidase with inhibitor in binding cavity

Author: K. Kaszuba, T. Róg, J.-F. St. Pierre et al.

Publication: Sar & Qsar in Environmental Research

Publisher: Taylor & Francis

Date: Oct 1, 2009

Copyright © 2009 Taylor & Francis

Logged in as:
sepideh soltani
Account #:
3000479427

LOGOUT

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

BACK | CLOSE WINDOW