MODELING AND SIMULATION OF FIBRINOGEN AND ITS ADSORPTION BEHAVIOR

Dissertation
zur Erlangung des Grades
"Doktor
der Naturwissenschaften"
am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

Stephan Köhler

geb. in Sonneberg (Thüringen)
Mainz, den 16. Dezember 2014

THESIS COMITEE:

DATE OF EXAM: June 15, 2015

The protein model resulting from the X - Ray crystallographic observations is a platonic protein, well removed in its perfection from the kicking and screaming stochastic molecule that we infer must exist in solution.

– G. Weber [1]

ZUSAMMENFASSUNG

In dieser Arbeit werden verschiedene Zugänge zur Modellierung und Simulation des Blutproteins Fibrinogen vorgestellt. Die verwendeten Methoden sind darauf ausgelegt systematisch verschiedene Zeit- und Längenskalen der Proteindynamik miteinander zu verknüpfen und das Verhalten von Fibrinogen in Lösung und an anorganischen Oberflächen zu erklären.

Als erstes werden atomistische Simulationen des Fibrinogen-Protomers und -Dimers präsentiert. Diese Simulationen wurden in explizit behandeltem Lösungsmittel durchgeführt um die Dynamik von Fibrinogen in Lösung zu verstehen. Eine unerwartet starke und schnelle Biegebewegung, welche durch Molekulare Gelenke ermöglicht wird, wurde identifiziert und über Winkelverteilungen charakterisiert. Dank der atomistischen Details dieser Simulationen ist es mögliche auch Effekte auf kleinen Skalen zu untersuchen. Als Beispiel wird ein möglicher allosterischer Effekt der Ligandenbindungstaschen untersucht. Weiterhin wurden die Adsorption von Fibrinogen an Graphit- und Glimmeroberflächen untersucht. Die Simulationen zeigen verschiedene Adsorptionsmechanismen an der hydrophoben Graphitoberfläche und der hydrophilen Glimmeroberfläche. Im Fall von Glimmer wird eine bevorzugte Adsoprtionorientierung identifiziert.

In vielen praktischen Anwendungen spielen Aggregate von Fibrinogen eine Rolle die nicht auf der atomistischen Skala simuliert werden können. Es ist deshalb nötig vergröberte Modelle zu entwickeln, welche die wichtigen Aspekte der atomistische Dynamik korrekt wiedergeben. Zwei solche Modelle werden im zweiten Teil der Arbeit präsentiert. Das erste Modell wird systematisch aus den atomistischen Simulationen abgeleitet und verringert die Zahl der zu simulierenden Teilchen von 31.000 Atomen auf 45 Kugeln. Die intramolekulare Wechselwirkung dieser Kugeln wird durch ein heterogenes Federnetzwerk wiedergegeben, während die intermolekularen Wechselwirkungen als Kombination von elektrostatischen und van der Waals Wechselwirkung ausgedrückt werden. Eine Methode zur Bestimmung der vergröberten Ladungsverteilung wird entwickelt. Das zweite Modell ist ein phänomenologisches Modell das aus harten Kugeln und Stäben besteht, die über Gelenke miteinander verbunden sind. Auf Grund seiner Einfachheit kann dieses Modell nur das Verhalten von Fibrinogen auf großen Skalen wiedergeben. Es erlaubt es jedoch experimentelle Ergebnisse zu Fibrinogenkonformationen auf anorganischen Oberflächen zu erklären.

ABSTRACT

In this thesis different approaches for the modeling and simulation of the blood protein fibrinogen are presented. The approaches are meant to systematically connect the multiple time and length scales involved in the dynamics of fibrinogen in solution and at inorganic surfaces.

The first part of the thesis will cover simulations of fibrinogen on an all atom level. Simulations of the fibrinogen protomer and dimer are performed in explicit solvent to characterize the dynamics of fibrinogen in solution. These simulations reveal an unexpectedly large and fast bending motion that is facilitated by molecular hinges located in the coiled-coil region of fibrinogen. This behavior is characterized by a bending and a dihedral angle and the distribution of these angles is measured. As a consequence of the atomistic detail of the simulations it is possible to illuminate small scale behavior in the binding pockets of fibrinogen that hints at a previously unknown allosteric effect. In a second step atomistic simulations of the fibrinogen protomer are performed at graphite and mica surfaces to investigate initial adsorption stages. These simulations highlight the different adsorption mechanisms at the hydrophobic graphite surface and the charged, hydrophilic mica surface. It is found that the initial adsorption happens in a preferred orientation on mica.

Many effects of practical interest involve aggregates of many fibrinogen molecules. To investigate such systems, time and length scales need to be simulated that are not attainable in atomistic simulations. It is therefore necessary to develop lower resolution models of fibrinogen. This is done in the second part of the thesis. First a systematically coarse grained model is derived and parametrized based on the atomistic simulations of the first part. In this model the fibrinogen molecule is represented by 45 beads instead of nearly 31,000 atoms. The intra-molecular interactions of the beads are modeled as a heterogeneous elastic network while inter-molecular interactions are assumed to be a combination of electrostatic and van der Waals interaction. A method is presented that determines the charges assigned to beads by matching the electrostatic potential in the atomistic simulation. Lastly a phenomenological model is developed that represents fibrinogen by five beads connected by rigid rods with two hinges. This model only captures the large scale dynamics in the atomistic simulations but can shed light on experimental observations of fibrinogen conformations at inorganic surfaces.

CONTENTS

Part I

INTRODUCTION

1

OVERVIEW

This thesis will use computer simulations and modeling techniques to investigate the behavior of the blood protein fibrinogen in solution and at inorganic surfaces. As a main component of vertebrate blood clots, fibrinogen is a protein of immense medical importance. Mutations of fibrinogen can influence the behavior and function of fibrinogen and lead to bleeding or immune disorders of varying degree. At the same time, the interaction of fibrinogen with biomedical materials can influence their biocompatibility by regulating the immune response or the risk of thrombus formation. Thrombi form if the aggregation of fibrinogen at the surface initiates blood clotting. This might lead to the blocking of blood vessels or the formation of a mobile thrombus that can lead to a thromboembolism or a stroke.

The importance of this topic has prompted interest in the research into the structure and behavior of fibrinogen. At this point, investigations of its molecular structure have been ongoing for more than 60 years. A lot of data has been collected experimentally with techniques ranging from electron and atomic force microscopy over X-ray crystallography to micro-calorimetry, hydrodynamic measurements, and biological activity assays.
Many of the structural features have, by now, been identified and genomic analyses have shed light on the evolution of fibrinogen. However, many of the mechanisms remain elusive in experiments due to the size of fibrinogen. Fibrinogen is a very large biomolecule of 340kDa. This makes it easy to pick out large scale structural features but obscures mechanisms on a smaller scale. Additionally, the size and complexity of fibrinogen complicates mutational studies.

For this reason modeling of fibrinogen behavior has always played a major role in the interpretation of experimental results and was key in the identification of its overall structural features and polymerization behavior. Thanks to high resolution structural data computer simulations can now provide detailed insights into the inner workings of proteins as big as fibrinogen. Over the last two decades computational techniques have advanced sufficiently to handle the large system sizes required to study proteins. The development of potentials to describe the interactions of protein atoms with each other as well as surfaces was another key to obtain insight into the atomistic dynamics of proteins. While the potentials are not perfect they have been

improved over the last years. A discussion of the strength and weaknesses of the relevant potentials is given in Chapter 3.

Knowledge about the atomistic dynamics are important but proteins play many of their key roles on time and lengths scales much larger than those accessible in all atom simulations. It is common to use phenomenological models of the studied system to remedy this. Such models can be used to extrapolate results from atomistic simulations or perform simulations of a simplified model. However, the drawback is that such models take only the key features into account that strike the modeler as important. This might miss relevant mechanisms in the complex dynamics of proteins. In recent years, there has been a push to systematically develop models of proteins based on atomistic simulation and thus eliminate modeling biases, at least to some extent. Such models systematically coarse grain the atomistic dynamics by reducing the simulated degrees of freedom. In principle, these approaches allow the accurate modeling and simulation of a system on multiple scales.

Blood clotting as well as the adsorption and layer formation of fibrinogen at surfaces are inherently multi scale problems. On the one hand the association of fibrinogen molecules with each other and the surface is driven by atomistic interactions and require great chemical specificity on a sub-nanometer scale. On the other hand the resulting aggregates can span micro- or millimeters and form on time scales of seconds or hours. While fibrinogen has been modeled before, no computational investigation of its solution behavior has yet been attempted. Accordingly, the current fibrinogen models are either phenomenological or describe only specific features of fibrinogen. In this thesis, the solution and adsorption behavior of fibrinogen is characterized for the first time on an atomistic level. Based on the results of these simulation the most prominent phenomenological fibrinogen model is improved and a new multiscale model is developed.

The thesis is divided into three parts. In the introduction, more details about fibrinogen and computational techniques for the study of multiscale phenomena are presented. In Chapter 2, the biochemistry, structure, and evolution of fibrinogen is presented first. After these more general remarks, the experimental literature on fibrinogen behavior is reviewed with a special focus on adsorption experiments Chapter 3 introduces the concepts of biomolecular simulations as they are used in this thesis. Special attention is paid to the sampling problem arising for multiscale processes and how to overcome it. The literature of fibrinogen simulations is reviewed as well as the literature on protein adsorption simulations.

The second part of the thesis describes the atomistic simulations of fibrinogen. The details of the molecular dynamics simulations of fibrinogen are described in Chapter 4 together with the

analysis techniques that were used to extract information from the simulations. The results of the solution simulations are discussed in Chapter 5. In these simulations, a molecular hinge is identified that gives the fibrinogen molecule its flexibility. The effect of sugar groups on the fibrinogen dynamics is presented as well as the dynamics of the polymerization sites. The latter show a signature of a potential allosteric effect. Chapter 6 describes the results of adsorption simulation at the mica and graphite surfaces. These simulations demonstrate that fibrinogen stays flexible even in the neighborhood of a surface. While the hydrophobic graphite is shown to be stickier than the hydrophilic mica surface, it is shown that fibrinogen adsorption is more specific on mica. The specific interaction between fibrinogen and mica gives rise to an asymmetry in the adsorption events.

The last part of the thesis is dedicated to the modeling of fibrinogen based on the atomistic simulation presented in the second part. Chapter 7 is dedicated to the systematic development of a coarse grained model of fibrinogen based on the atomistic simulations in solution. Different approaches to coarse grained protein models are briefly reviewed before the fibrinogen model is developed. The model structure is determined based on correlated motions in fibrinogen while the dynamics are reproduced by an elastic network. First, steps toward the parametrization of intermolecular interactions are shown. An improved phenomenological model is presented and analyzed in Chapter 8. This model captures the large flexibility of fibrinogen and can explain experimentally observed differences in its adsorption behavior.

The results are summarized in Chapter 9 and future avenues of research are outlined.

2

FIBRINOGEN

Fibrinogen is a large (340 kDa) soluble fibrous glycoprotein occurring in the blood of vertebrates. After an injury to a blood vessel, fibrinogen forms fibers that assemble into a network and form the basis for a blood clot [2]. Fibrinogen is involved in haemoragic disorders and the activation of fibrinogen can occur when blood comes in contact with medical equipment, such as stents or dialysis machines. Therefore, it is of great interest to understand the behavior of fibrinogen in solution and at surfaces. In the following chapter, a short introduction into the biochemistry of proteins is given before the characteristics of fibrinogen and its known behavior are described. The role of fibrinogen in the clotting mechanism of vertebrates and its evolution are discussed. The chapter closes with a review of what is currently known about the behavior of fibrinogen at surfaces both from experimental and theoretical studies.

2.1 BIOCHEMISTRY OF FIBRINOGEN

2.1.1 *Protein Structure: An Overview*

Together with DNA, RNA, and lipids, proteins form the basic molecular components of life on earth. While DNA mostly stores information, proteins have a myriad of functions in organisms. This variety of functions is addressed by a variety of forms that proteins take. A main tenet of molecular biophysics is that the structure and function of a protein are related. A fact that can be illustrated by destroying the structure of a protein and observing its loss of function. Typical ways to destroy the structure of a protein are temperature or pH changes.

At the heart of protein structure is a hierarchical organization of different structural levels. In the following the different levels of protein structure are described briefly. A more detailed overview can be found in textbooks, e.g. [3]. Protein structures that have been experimentally determined are available in the Protein Data Bank (PDB) [4] where they are identified by a four letter alphanumeric code (PDBID).

Primary Structure

A protein is a linear polypeptide and its primary structure is the sequence of amino acids that it contains. In the following amino acids will also be referred to as residues. An amino acid consists of a carbon atom with a primary amine, a carboxylic acid group, and a side chain. This carbon is called the C_α carbon. There are twenty different naturally occurring side chains giving rise to twenty amino acids that make up all proteins. The peptide bond is formed by a condensation reaction between the amine group of residue N and the carboxylic acid group of residue $N+1$. The atoms not belonging to the side chain thus form the backbone of the protein. The residues in a protein are counted from the amino terminal (N-terminus) to the carboxyl terminus (C-terminus).

Each of the twenty amino acids is coded for by a sequence of three DNA base pairs. To synthesize the protein, the DNA is first translated to RNA and then transfered to the ribosome where the synthesis (translation from RNA to protein) happens. Because there are more than twenty three letter combinations that one can form with four bases, each amino acid is coded for by multiple three letter codes. Additional codes exist for the start and end of a protein sequence as well as the regulation of translation. An example of the primary structure of a short peptide is shown in Fig. 1.

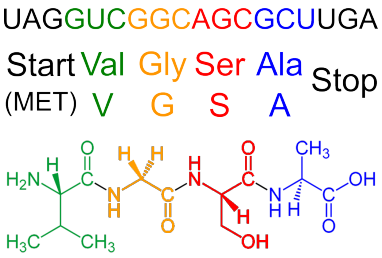

Figure 1: DNA sequence, amino acid sequence and chemical structure of a short peptide. The start codon is always translated as a methionine and is not shown in the peptide structure.

Among the naturally occurring amino acids proline can be singled out as the only amino acid where the side chain attaches back to the backbone amine group (see Fig. 2). The ring formed in this fashion introduces constraints on the peptide bonds and forces certain torsion angles along the backbone. On the other hand the side chain of glycine (see Fig. 1) is just a single hydrogen atom and imposes no constraints. Other amino acids have backbone flexibilities between those of glycine and proline. The side chain also influences other properties and amino acids are usually grouped into charged, polar uncharged and hydrophobic categories according to their side chain.

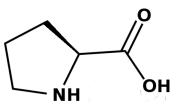

Figure 2: The structure of proline forms a ring along the backbone which increases its rigidity.

For specialized function a protein can undergo post translational modifications after its synthesis. Such modifications include the formation of disulfide bonds between cysteine residues or the covalent attachment of sugar groups, so called glycosilation. The formation of disulfide bonds is especially important in harsh or unpredictable environments such as the extracellular medium as the the additional covalent bonds stabilize the structure. Sugar groups can also stabilize a protein by crowding effects and the formation of hydrogen bonds [5, 6]. Glycosilation however does more than stabilize protein structure. The sugar groups are an important factor in protein-protein as well as protein-cell interactions. It is thus not surprising that glycosilation patterns are often evolutionarily highly conserved [7].

Secondary Structure

The first structural level in three dimensions is the so called secondary structure. The secondary structure arises from the torsional degrees of freedom around the C_α–N and C_α–C bonds in the protein backbone. To clearly visualize the secondary structure, often only the path of the backbone atoms is traced (see Fig. 3)

The main secondary structure elements found in proteins are α-helices and β-sheets. Both of those structures are stabilized by hydrogen bonds between carbonyl and secondary amine groups in the backbone. A β-sheet consists of multiple parallel strands between which hydrogen bonds are formed. In the α-helix the hydrogen bonds form between the amine of residue N and the carbonyl of residue $N+4$. The alpha helix turns thus have a periodicity of 3.6 residues along the sequence. Beside the α-helix, other helix types exist. They are characterized by a different hydrogen bonding pattern or a specific sequence such as polyproline or polyglycine helices.

As was already mentioned, the favorable torsion angles depend on the side chain of the amino acid. Certain amino acids preferably form certain structures, e.g. the polyalanine peptide forms α-helices while proline disrupts them. Proline disrupts helices because its constrained backbone conformations are not compatible with the turn of an α-helix. On the opposite end of the flexibility range glycine can disrupt helices precisely because it is too flexible and the side chain does not constrain the backbone enough. Predicting the three dimensional structure of a protein based on its sequence is the problem of protein folding and has attracted a lot of interest both experimentally and theoretically due to its importance for diseases and potential industrial applications [8–11]. The folded state of a protein is termed its native state. A protein that has lost its native conformation is called denatured. The myriad interactions and sterical restrictions in proteins lead to a very rugged energy landscape not unlike those of glasses [12, 13]. Just like in glasses the complex structure of this free energy landscape can lead to a protein becoming trapped in a metastable state surrounded by high free energy barriers. In that case the protein can not reach its native state on physiological time scales. The resulting protein conformation is said to be misfolded.

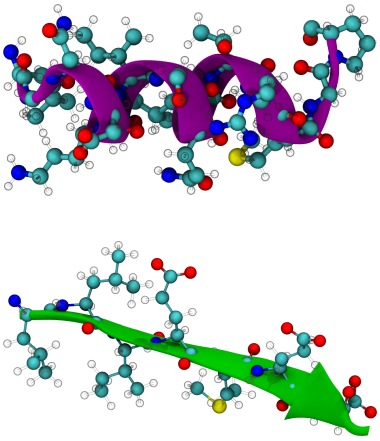

Figure 3: The α-helix (top) and a strand of a β-sheet (bottom) are two of the most common secondary structure elements in proteins and are often represented as cartoons along the backbone.

Besides these repetitive secondary structure elements, irregular elements exist. These elements form turns or coils/loops and can also be stabilized by hydrogen bonds to form rigid structures. Such loops can form binding sites for metal ions such as, e.g. Ca^{2+} in extracellular proteins. Metal ions can be bound in predefined binding sites or can help stabilize the protein structure as is the case for the so called zinc finger. Loops that are not stabilized by hydrogen bonds are usually mobile, disordered regions.

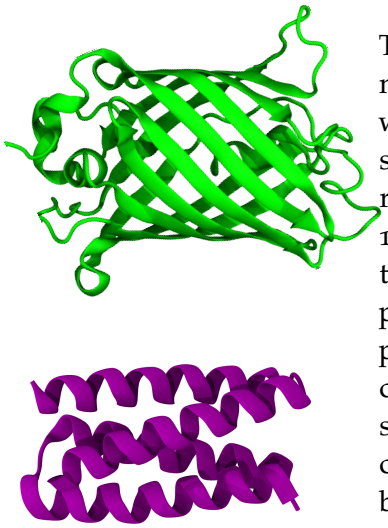

Figure 4: Examples of a β-Barrel in Green Fluorescent Protein (top, PDB ID 1CV7) and a 4-α-helix bundle (bottom).

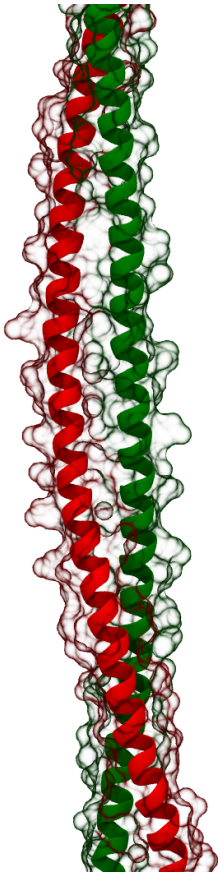

Figure 5: Keratin structure (PDBID: 3TNU) showing a coiled-coil sequence of two α-helical strands.

Higher Order Structures

The third level in the protein structure hierarchy is the aptly named tertiary structure. This term refers to the folding of a whole protein domain. A protein domain is part of the protein sequence that will fold into the correct conformation even if the rest of the protein is missing. Such domains are typically 50 to 100 residues in size and are evolutionarily mobile. This means that similar domain structures can be found in non homologous proteins [14]. Domains of extracellular proteins are more tightly packed than domains of intracellular proteins. In addition extracellular domains show ligand binding sites more frequently, are stabilized by disulfide bonds and regulated by the presence of calcium ions. Typical examples of tertiary structures are α-helix bundles or β-barrels as illustrated in Fig. 4.

The last structural level is the quaternary structure which describes the arrangement of domains. These domains may belong to the same peptide chain or be part of a multi-protein complex. Such complexes can be held together by non-covalent forces or disulfide bonds.

Based on the structural organization of proteins, they can be classified into globular, fibrous or intrinsically disordered proteins. Globular proteins have a compact conformation where several secondary structure elements are arranged to form the domain. Fibrous proteins are dominated by helical secondary structures that form highly elongated molecules. Fibrous proteins often play structural roles in organisms as is illustrated by collagen and keratin, two of the most prominent proteins in this class. To perform these structural functions, fibrous proteins assemble into insoluble fibers. The functions of globular proteins are much more diverse, ranging from enzymes to transport proteins, and proteins for osmotic regulation. In contrast to globular and fibrous proteins, intrinsically disordered proteins do not adapt a native structure with stable secondary structure elements.

Coiled-Coil Structures

For the study of fibrinogen it is important to recall the defining features of fibrous proteins: the so called coiled-coil regions. An example is shown in Fig. 5. In these structures different peptide strands of helical conformation are twisted around each other to form a coiled -coil motive. The coiled-coil motive consists usually of two or three helical strands, but proteins with structures involving four or more helical strands exist. Increasing the number of helical strands can either increase the stability by forming a larger hydrophobic core or it can form tubes that can be filled by water or other molecules [15].

To form the association necessary to keep a coiled-coil structure stable, each helix has a hydrophobic strip along its contour that promotes the lateral association of helices. This requires a

certain regularity with which different types of amino acids occur in the sequence. The spacing of hydrophobic residues along the sequence determines if the hydrophobic strip runs straight along the helix or if it winds around it. Only if the strip winds around it a twisted coiled-coil region is formed. Depending on the length of the segment and the regularity of the amino acid sequence coiled-coil motives can be grouped into one of five categories A–E [16]. Obviously the spacing also determines if a left or right handed twist results. Futhermore, the packing in coiled-coil regions is much more regular than in the associations of helices in globular domains. This is facilitated by a knob-into-holes mechanism where the side chain of a residue from one helix is surrounded by four side chains from the second helix [15].

Natural coiled-coil structures are frequently not perfect and display defects. If three or four residues are inserted, the disturbance is close enough to the α-helical periodicity so that they can be accommodated by a slight distortion of the ideal structure. Larger distortions are extruded as loop regions to preserve the coiled-coil. Such extrusions are particularly favored by the presence of glycine at their beginning and end.

2.1.2 *The Structure and Function of Fibrinogen*

Due to its physiological relevance, fibrinogen has long been of interest to researchers. The first investigation into the structure of fibrinogen dates back to the dawn of protein structure research in the late 1940s and 1950s. Even the low resolution, offered by the first electron microscopy techniques showed that fibrinogen consists of three major globular regions that are connected by two thin connector regions [18]. It took more than 50 years of effort until an atomic resolution crystal structure of fibrinogen was determined [19] that filled in most of the details. A schematic of the fibrinogen structure is shown in Fig. 6. In the following, the identified structure of fibrinogen and what has been learned about its behavior are summarized.

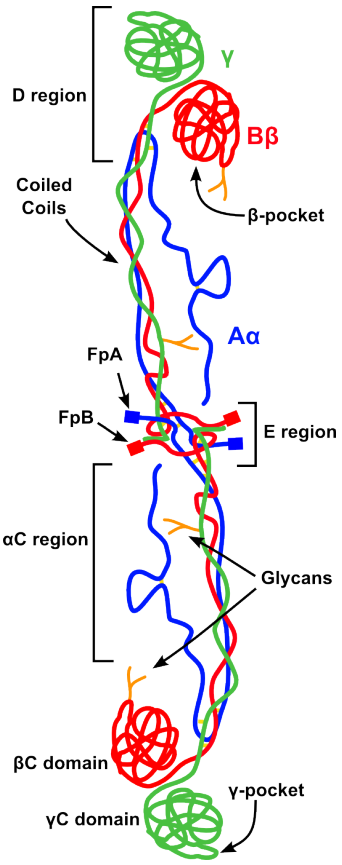

Figure 6: Schematic drawing of fibrinogen (after [17]) showing the different chains, glycosilation and polymerization sites.

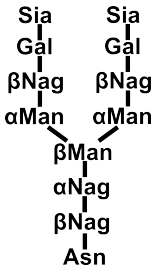

Figure 7: The structure of the asparagine (Asn) linked biantennary carbohydrate clusters in fibrinogen. Abbreviations: α/βNag: N-Acetyl-α/β-D-glucosamine, α/βMAN: α/β-D-mannopyranose, Gal: β-D-Galactopyranose, Sia: 5-N-acetyl-α-D-neuraminic acid.

Fibrinogen Structure

Fibrinogen consists of three pairs of identical peptide chains α, β, γ. All three chains start with N-terminal regions without regular secondary structure elements followed by long α-helical regions. In the fibrinogen molecule, these helical segments form a coiled-coil region. In contrast to the α- and β-chains, the helical region in the γ-chain is broken in the middle of the helical section. This break is caused by the presence of two proline residues.

The α-chain has a turn region followed by another α-helix and ends with a largely disordered part called the αC-region [20]. The αC-region is not completely disordered and a short β-hairpin structure stabilized by a disulfide bond has been identified [21, 22].

Figure 8: X-ray crystal structure of human fibrinogen. The flexible N-terminal parts, the αC-regions and the carbohydrate clusters in the coiled-coil region have not been resolved.

Both the β- and γ-chains have globular domains following their helical segments. Those globular domains are called βC- and γC-domains respectively [2]. They can fold independently from each other into functional forms, but show increased stability after they associate [23]. Furthermore, both βC- and γC-domain contain ligand binding sites. The β-pocket is located between a carbohydrate cluster and the coiled-coil region, while the γ-pocket is located at the outer end of the γC-domain. These binding pockets are regulated by binding sites for Ca^{2+} ions [24]. Other binding sites in the D-domain are responsible for interaction with integrins. Integrins are trans-membrane signaling proteins in the cell membrane.

As was already alluded to, the β- and γ-chain have asparagine linked carbohydrate clusters attached to them. The carbohydrate clusters have a biantennary structure which is given in Fig. 7 and greatly increases the solubility of fibrinogen. In the β-chain, the glycosilation site is in the globular domain while it is in the coiled-coil region for the γ-chain. The globular fibrinogen domains have thus all the hallmarks of typical extracellular protein domains. Indeed, fibrinogen like domains and coiled-coil regions belong to very widely spread families of protein domains reaching far beyond blood proteins [14].

The three distinct chains form a fibrinogen protomer by assembling along their helical segments in a two step process. First,two chains assemble along their long helical segments and then the third joins to form a coiled-coil region [25]. The result is a type C coiled-coil region, i.e. one consisting of short stacked fragments that have only poor conservation of the amino acids in individual positions [15]. This structure is stabilized by disulfide rings at the N- and C-terminal ends [25]. Further disulfide bonds stabilize the βC- and γC-domains that together form the D-region. The globular domains are not symmetric with respect to the coiled-coil region, but sit on top of it. In the assembled protomer, the α-chain folds back onto the coiled-coil region forming a short segment of a four stranded coiled-coil. The end of the αC-region is electrostatically bound to the N-terminal part of the protomer under physiological conditions [26], but its length and flexibility nevertheless influence the hydrodynamics behavior of fibrinogen [27].

Two protomers are linked together by disulfide bonds at their N-terminal ends to form a fibrinogen molecule. After the dimer is formed, no free cysteines remain. The central region of fibrinogen is called the E-region and forms a globular region with limited secondary structure. The interaction between two protomers is not only ensured by the disulfide bonds but also non-covalent interactions. The β-chain of one protomer forms a two stranded β-sheet that is threaded through a loop formed by the α-chain of the opposing protomer and forms a "molecular handshake" between the protomers [28]. The two protomers are linked together in such a fashion that they are rotated by 180° around the coiled-

coil region with respect to each other. The N-terminal parts of the α- and β-chains extend significantly beyond the disulfide bonds in the E-region and contain short peptide sequences that can be cleaved by the enzyme thrombin. The released peptides are called Fibrinopeptides and labeled according the chain from which they came. To indicate the presence of the fibrinopeptides, the chains are written Aα and Bβ. This cleaving exposes binding sites, so called knobs, that are inserted into the pockets in the D-region [24, 29]. This process leads to fibrinogen polymerization and is described in more detail in section 2.1.3. The B knob is especially flexible sitting at the end of a 50 residue long tether beyond the E-region. This is necessary as the β-pocket is far less accessible than the γ-pocket. The structural features thus described are summarized in Fig. 6. Because of the flexibility of the αC-region and its lack of extended structures, it is often not possible to resolve it and studies have to focus on the more rigid fibrinogen core, that is the fibrinogen dimer without the αC-region.

These structural consideration have been derived over the years from X-ray structures of parts of the fibrinogen molecule and except for the αC-region they have been confirmed by an atomic resolution X-ray structure [19] that is shown in Fig. 8.

Fibrinogen physiology and function

The three fibrinogen chains are encoded by three different genes called FGA, FGB, and FGG and located on chromosome 4 [30]. The FGB and FGG genes are regulated together but separately from the FGA gene. *In vivo* a similar expression level of all three chains is the result of both regulation mechanisms being coupled to the plasma fibrinogen levels. In healthy humans alternative splicing variants of the FGA and FGG genes exist. The alternative variant of the FGA gene gives rise to an α-chain that is extended at the C-terminus and ends in a globular domain that is a homologue of the βC- and γC-domains. This variant makes up 1–2% of fibrinogen molecules and its function is still unknown. The FGG variant results in a γ-chain that is extended by 20 residues at the C-terminus. Such a modified γ chain occurs in roughly 15% of fibrinogen molecules and seems to play a role in the regulation of thrombin activity [31].

In humans, the fibrinogen molecule just described is synthesized mainly in the liver at rates of (1.7–5.0)g/day. In the case of injury, this production rate can be increased. Due to the complex nature of fibrinogen, problems can occur during the synthesis. Alternatively mutations of the genes might code for non-functional version of the protein. To avoid problems from missformed fibrinogen entering the blood stream, control mechanisms are in place that will prevent the excretion of non-functional fibrinogen to a large degree. For example, this is the case if the C-terminal disulfide ring is not formed correctly. Other cases where mutations lead to non-functioning protein are discussed

in Chapter 5. In healthy individuals, non-functional fibrinogen is degraded within the cell.

Most of the synthesized, functional fibrinogen is excreted into the blood where it remains for three to five days before it is degraded by plasmin. The resulting concentration of fibrinogen in human blood is 2.5g/L which makes it the fifth most common protein in blood plasma [31]. The lifetime of fibrinogen is rather short compared to other plasma proteins like albumin that remain in the blood for up to three weeks. Only 2–3% of fibrinogen are used up in blood coagulation in healthy individuals before the protein is degraded. The degradation occurs in a two step process. In the first step, plasmin removes the flexible αC-regions and N-terminal residues. In the second step, it cuts the middle of the coiled-coil region, where the unfolded part of the γ-chain serves as an attack site [31–35]. The residues at which plasmin cuts bonds are given in Table 1.

Chain	*Residues*
α	*16–17, 19–20, 78–79, 81–82, 104–105, 110–111, 199–200, 205–206, 219–220, 230–231, 239–240*
β	*14–15, 42–43, 53–54, 121–122, 133–134*
γ	*53–54, 58–59, 62–63, 85–86, 88–89, 405–406*

Table 1: Attack sites of plasmin in fibrinogen [25, 35–38].

In the blood, the main function of fibrinogen is blood clotting. To this end, it is converted by thrombin to fibrin in which form it can polymerize and form fibers. This process is described in more detail in the next section. Due to its importance in the clotting process, it has been shown that high fibrinogen levels are markers for, but not contributors to, cardiovascular disease [39]. Low fibrinogen levels on the other hand can cause significant problems. Due to mutations or metabolic problems, the fibrinogen concentration in the blood can be reduced (hypofibrinogenenmia) or completely absent (afibrinogenemia). In hypofibrinogemic patients, blood clotting is compromised and haemophilia like symptoms are observed. These condition usually occur in people with heterozygous mutation for dysfunctional fibrinogen. If fibrinogen is present in the blood in a non-functional state, the condition is called dysfibrinogenemia.

Fibrinogen does not only bind to itself but also to many other proteins and is an important part of the regulatory network of extracellular proteins. The proteins regulated by fibrinogen are related to plasmin action, the clotting cascade, and growth factor and cell regulation [40–44]. As a consequence, growth factors and immune response can be localized to the site of injury. The clotting cascade and the involved proteins are described in more detail in section 2.2. Not for all of them are the binding sites known to date and it is often the case that the binding sites are cryptic in fibrinogen but become exposed in fibrin [2].

The main interaction with cells occur in the fibrin form. One example is the the exposure of a heparin binding site in the E-region that allows fibrinogen interactions with various other cell types, it can, e.g. stimulate platelet spreading or endothelial cell proliferation. The two cell types fibrinogen mostly interacts with are platelets and leukocytes. Fibrinogen has binding sites for platelet integrin $\alpha_{II}\beta_3$ in the γC-domain. This allows platelets to attach to the fibrin fiber network and seal a wound. It also enables fibrinogen to link two platelets by attaching each of the

γC-domains to a different platelet. With this mechanism, blood clots can be strengthened. Other integrin binding site in the γC-domain highlight fibrinogens role in the immune system. To fulfill its function in the immune system, less fibrinogen is needed than for blood clotting. For this reason even patients with hypofibrinogenemia can have a functioning immune system [30].

The integrin binding sites can bind leukocyte integrin $\alpha_M\beta_2$ and allow the attachment of these cells to a fibrin matrix. The behavior of leukocytes can be regulated by mechanosensing for which the fibrin matrix provides the substrate [45, 46]. The P1 leukocyte integrin binding site is cryptic in fibrinogen in solution but can become exposed upon adsorption and fibrin formation. The P2 site forms a β-sheet with P1, but can be displaced without disrupting the rest of the γC-domain. The residues forming the various known binding sites for proteins are listed in Table 2.

As a consequence of the plethora of interactions of fibrinogen with other molecules, mutations can significantly impact the health of affected patients. Mutations of fibrinogen are known to lead to bleeding disorders like pulmonary embolisms [47], thromboses [48], or complications in pregnancy [49]. A lack of glycosilation affects the chaperone activity fibrinogen has for plasma transthyretin. A chaperone protein is a protein that helps other proteins fold into their native state by presenting favorable interaction sites. If the chaperone activity of fibrinogen is influenced, transthyretin has an increased chance to fold into the wrong conformation. This misfolding leads to amyloid fibril formation that in turn leads to amyloidotic polyneuropathy, i.e. a damage of the nervous system [50]. There are also known mutations of fibrinogen that lead to renal amyloid disease [51].

Besides these interactions with proteins of the same organism, fibrinogen can interact with bacterial proteins [3], a property that is also found in many fibrinogen related domains [31]. An example of this are bacterial proteins from streptococcus. With the help of these proteins, fibrinogen forms a network that is efficient in activating an immune response against the bacteria [54].

Molecule/Fragment	*Binding Site*
	$\gamma 400$–$\gamma 411$
Plattlet Integrin $\alpha_{II}\beta_3$	$\alpha 572$–$\alpha 575$
	$\alpha 95$–$\alpha 98$
Leukocyte Integrin $\alpha_M\beta_2$	$\gamma 190$–$\gamma 202$
	$\gamma 377$–$\gamma 395$
$\alpha_5\beta_1$	$\gamma 365$–$\gamma 383$
RIBS	$\gamma 373$–$\gamma 385$
Heparin	$\beta 15$–$\beta 57$
Cadherin	$\beta 15$–$\beta 42$

Table 2: Known binding sites in the fibrinogen molecule [40–44, 52, 53]. The binding sites for the leukocyte integrin $\alpha_M\beta_2$ and $\alpha_5\beta_1$ are referred to as P1–P3.

2.1.3 *Fibrin*

In the previous section, the formation of fibrin from fibrinogen was already alluded to and will now be described in detail. The first step in fibrin formation is the transformation of fibrinogen to fibrin monomers. This is accomplished by the enzyme thrombin that cleaves specific bonds in the N-termini of the Aα and Bβ regions, thereby releasing the fibrinopeptides. This is not a simple single step process. While thrombin can cleave both the FpA and FpB bonds at any time, the release of FpB is only substantial after FpA has been released and self assembly has begun. The cleavage of the fibrinopeptides exposes two sections termed the A- and B-knob. The activation of a protein by removing parts

of it is called limited proteolysis and is one of the dominant regulatory mechanisms in the extracellular environment.

The sequence of the A-knob binding in the γ-pocket is a glycine-proline-arginine-proline sequence, while it is the glycine-histidine-argininie-proline stretch that forms the interaction between the B-knob and the β-pocket [55]. The binding of these peptides is very specific. While both pockets can bind each peptide if the other is not present, they exclusively bind the one just mentioned if both are present [24, 56, 57] despite just one residue difference. It should be pointed out that the β-pocket is the more promiscuous binder of the two pockets and can also accommodate peptides starting with an alanine instead of a glycine residue. This is not possible in the γ-pocket [56].

The self assembly process from fibrin monomers to fibrin fibers is in a first step driven by electrostatics. The fibrinopeptide regions are highly charged, this is how they are identified by thrombin [2], and their release changes the E-region charge by 13e. This change in electrostatic interaction leads to an attraction between the E-region of one fibrin monomer and the D-region of two others. The electrostatic interaction also drives the insertion of the A-knob into the γ-pocket [55]. Thereby a half staggered, two molecules wide protofibril is formed and grows in length. This protofibril is depicted in Fig. 9 and its arrangement gives rise to a characteristic pattern of stripes with a 22.5nm periodicity in fibrin [2]. Due to the asymmetric positioning of the globular βC- and γC-domains, a twist is introduced into the protofibril so that longer structures will twist around themselves after 20 to 25 monomers [31, 58]. Unsurprisingly this step is inhibited if a synthetic peptide mimicking the A-knob is pre-bound to the fibrin monomers [59]. The formation of fibrin can also happen if only FpB is removed from the E-region. In this case, the interaction holding the fibers together are weaker and the fibers are only formed below 15°C. Similarly, the normal association behavior is not infinitly stable and no fibrin is formed above 41°C [60].

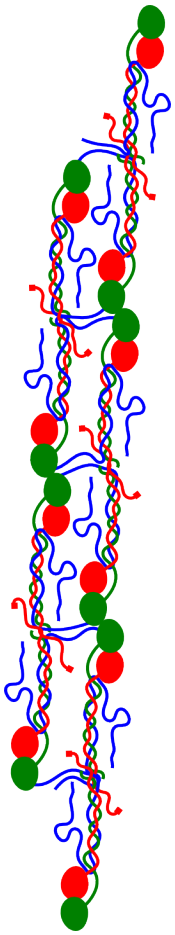

Figure 9: The fibrin protofibril forms after cleavage of fibrinopeptide A by electrostatic interactions between D- and E-regions. The result is a two molecule wide, half staggered arrangement of fibrin monomers.

The interactions between the peptide knobs and the pockets only serve to assemble the fiber. In this non-covalently bound state, a fibril would be of little use against the forces of flowing blood. Therefore, covalent links have to be established between different fibrin monomers. These covalent bonds between D-regions are introduced shortly after the protofibril is formed by the enzyme factor XIII which is active in a complex it forms with the fibrin monomers and thrombin at the junction [2]. The covalent bonds are formed between a lysine at γ406 and glutamines at γ398 and γ399 close to the C-terminus of the γ-chain. This cross-linking makes the protofibril much more resilient to stretching, allowing it to be stretched to 1.8 times its length and recovering [44]. The association of D- and E-region also exposes new thrombin binding sites that allow a more efficient cleavage of FpB [61].

The release of FpB allows the B-knob, which is located at a much longer, flexible tether than the A-knob to find its way to the β-pocket. This association stabilizes the protofibril further and deforms the βC-domain. The deformation results in a pulling away of the globular βC-domain from the coiled-coil region on top of which it sits in fibrinogen [62]. Due to the charged nature of the carbohydrate group and their position at the β-pocket, it is reasonable to assume that they play an important role in this assembly process [2, 19].

All four knobs in the E-domain are associated with the same two fibrin D-domains. Thus, a protofibril has no dangling knobs and is a self contained structure. This makes it identifiable as a distinct step in the fibrin formation process [63].

After the protofibril has grown to a sufficient length, lateral aggregation starts. Cleaving the FpB from the E-region also removes the electrostatic interactions that keep the αC-region attached there. As a result, the αC-regions can now move freely and induce interactions between neighboring protofibrils [2, 64]. These interaction happen mostly through interaction of the β-hairpin structures in these domains [65]. The resulting fibers form contacts along the globular domains in the D-regions [62] and are schematically shown in Fig. 10. The contacts in the globular regions are enabled by the deformation of the βC-domain during binding of the B-knob. An observation that is supported by the fact that fibrin monomers with pre-bound synthetic B-knob peptides form fibrin with enhanced lateral aggregation [56]. The fibers assembled in this way are cross-linked in the αC-region by factor XIII. Because the fibers are well formed at this point, the diffusion of factor XIII is hampered and it can take hours or days until the αC-cross-linking is complete [31]. The fact that factor XIII can diffuse through fibrin fibers already hints at the fact that the network is not as dense as one might think. And in fact fibrin constitutes only 0.25% of the volume of a clot while the rest is liquid [48].

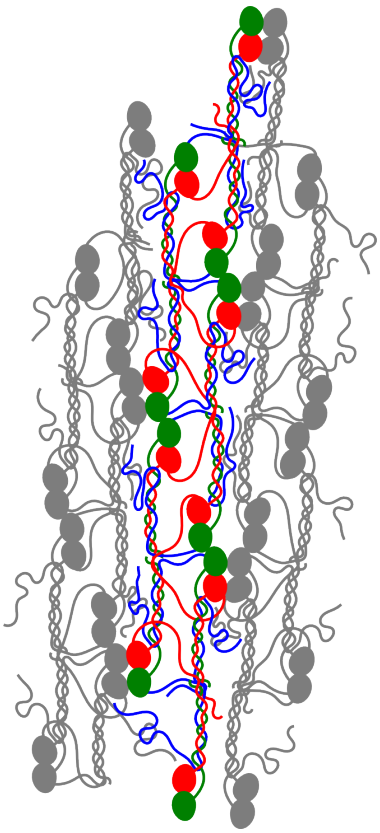

Figure 10: Approximate two dimensional representation of three protofibrils forming a fibrin fiber. In reality, the protofibrils and fibers are twisted around each other. The fiber is held together by cross-links in the αC-regions as well as non-covalent interaction in the γC- and βC-domains. These are enabled by deformations in the βC-domain upon binding of the B-knob.

The fibrin fibers assembled in this way are among the most extensible biopolymers found on earth and can be stretched to up to six times their length. Their complex nature also makes stretching experiments with atomic force microscopes or optical tweezers extremely hard to interpret as an observed signal could result from many different processes [66].

Because of the twist of individual protofibrils, the resulting fibers are also twisted. This results in somewhat compressed fibrin monomers in the core of the fiber while the molecules on the outside are stretched. This stretching provides a self regulation mechanism that limits the fiber diameter if the energy used for stretching is greater than the energy gained by binding [58]. The size at which the limiting point is reached depends on the physiological conditions and fiber diameters can vary between 50 and 200 monomer layers. The average fiber diameter is rather robust at 85nm under varying conditions, but the variation is increasing

for harsher conditions and for high pH and ion concentration a threshold is reached and only thin fibers are formed [60]. The deformation of fibrin monomers in the twisted fiber conformation can also expose binding sites that are not accessible in fibrinogen [31].

The twisting mechanism does not only limit fiber growth it also produces straight fibers. However, the fibers themselves are flexible and can diverge from one another. In this way, fibers can branch off of each other and form a space-filling network. The flexibility allows fibers in turn to bend and fuse together for part of their length [60]. The resulting fibrin network forms a viscoelastic gel that is the basis for blood clots. Depending on ion concentration, pH, and other proteins present, the release of the fibrinopeptides is modified so that the clot structure is variable in space and time [67]. *In vivo* the clot is covered by collagen and cells and forms the scaffold for new tissue.

Fibrin is very effective at forming a gel: already if 10% of the final protein is incorporated into the clot a space filling gel is obtained. This gel continues to increase its stiffness as more fibrinogen is incorporated and rearrangements take place. This process can take more than 24 hours and the result depends on the physiological conditions that influence the final network structure and which fibrinopeptides were cleaved. Typical values for the stiffness of fibrinogen are in the range of 0.1–600$\frac{\mathrm{N}}{\mathrm{m}^2}$. These values are constant under a wide variety of flow conditions. Only at high strain rates is the stiffness of the fibrin network affected. In an effect called strain-hardening, the stiffness increases at higher strain rates. Another effect that is not yet understood is the self healing ability of fibrin networks. Even after inelastic deformations over a long period of time, the mechanical properties of the gel are unchanged. One possible explanation for this behavior is the slippage of protofibrils in fibrin fibers [2]. It is worth noting that cross-linking with factor XIII almost entirely removes the viscous behavior of the fibrin gel. Without this cross-linking, fibrin is a very fast changing unit in which the fibrin monomers are exchanged freely and frequently [48]. Even if just the αC-regions are cross-linked, the elasticity of fibrinogen is increased [68].

A blood clot *in vivo* does not consist solely of fibrin and water. Instead, it contains many other proteins and cells, all of which change the network structure and mechanical properties. The viscoelastic properties of blood clots *in vivo* are of paramount importance as they determine if, e.g. a clot breaks loose and leads to a stroke.

2.2 EVOLUTION OF FIBRINOGEN AND THE CLOTTING CASCADE

2.2.1 *The Clotting Cascade*

The formation of fibrin just discussed is only the last step in a tremendous system of biochemical reactions that forms the so called clotting cascade and will be described with its evolution in the following.

The Clotting Cascade in Humans

The clotting cascade is a network of proteolytic reactions that involves nearly two dozen substances. The substances involved in the clotting cascade are called factors and assigned roman numerals. A deficiency in any of the clotting factors leads to various types of haemophilia. Most of the clotting factors are plasma glycoproteins that are synthesized in the liver and seven of these proteins are inactive proteases that become activated during the clotting process. These proteases are very specific, each cutting just one or a few bonds leaving the rest to downstream enzymes. An important clotting factor that is not a protein are Ca^{2+} ions that play a crucial role in regulating the activity of the proteins involved. In fact, decalcified blood is not able to clot. For a detailed overview of the different steps in the clotting cascade, the structure of proteins involved, and the proteolytic reactions the reader is referred to the literature [3, 31]. Here only a brief overview is given over key steps and a sketch of the clotting cascade is presented in Fig 11. All factors discussed in the following are proteins.

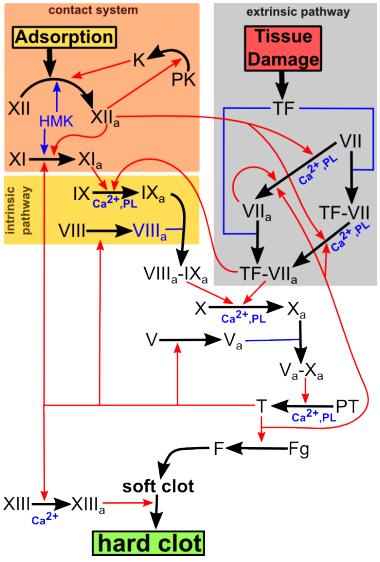

Figure 11: The network of reactions that forms the clotting cascade. Red arrows indicate proteolytic activation of a protein, blue are accessory factors. The activated state is denoted by the subscript a. Proteins not commonly associated with their clotting factor nomenclature are abbreviated as follows: (P)K: (pre)kallikrein, HMK: high molecular weight kininogen, PL: phospholipid, (P)T: (pro)thrombin, F: fibrin, Fg: fibrinogen, TF: tissue factor. Adapted from [3].

The main purpose of the different factors in the clotting cascade is to trigger a large response to a possibly small disturbance of blood vessels. This is achieved by many enzymes that only cleave specific bonds on substances that are more readily available than themselves, e.g. fibrinogen is about 50 times as abundant in blood as clotting factor X. Together with feedback mechanisms (see Fig. 11), this hugely amplifies the initial reaction. Such an amplification carries the risk of a runaway reaction that would lead to massive thrombus formation and be likely fatal. Hence mechanisms are in place to keep clotting in check, localize the clot to the site of injury, and dissolve clots after a certain time.

In humans, there are two major pathways that can trigger blood clotting. Both pathways are distinct ways in which factor X gets activated. This is an important step as factor X is the enzyme that converts prothrombin to thrombin which in turn leads to fibrin formation. Due to their complex nature, the two pathways are not really separate but interwoven so that they form a single clotting scheme.

The first pathway is called the extrinsic pathway. In this case, the blood is exposed to tissue factor, a protein in cell membranes

of various tissue types. Notably tissue factor is not present in the tissue lining the blood vessels and its presence is thus an indication of injury. The second pathway is called the intrinsic pathway and is initiated by the activation of certain clotting factors on negatively charged phospholipid membranes. Alternatively this pathway can be triggered by the deformation of clotting factors at surfaces, this specific manner is called contact pathway. For initiation of the contact pathway negatively charged surfaces are especially potent.

The clot is localized to the site of injury not only by dilution of the proteins associated with the clotting cascade but also by protease inhibitors that usually deactivate clotting factors within seconds. In addition to that, thrombin activates protein C which attacks other clotting factors and keeps the cascade in check. The activity of thrombin is tightly regulated by antithrombin and around 8 minutes after the clot appears thrombin activity rapidly drops. A clot is dissolved by plasmin. This protease is activated through tissue plasminogen activator (tPA) which in turn binds to fibrin, but not fibrinogen. The clot lysis starts in the coiled coil region of fibrin monomers where tPA, plasminogen and fibrin form a tertiary complex [2].

Evolution of the Clotting Cascade

The complex system that is the clotting cascade did not appear out of nowhere. While many organisms today have clotting mechanisms, the fibrin based clotting cascade is unique to vertebrates. Other clotting schemes evolved independently. The close relation between the clotting systems of different vertebrates is illustrated by the fact that blood from one vertebrate can clot blood from other vertebrates, albeit less effectively. It is even possible to clot human blood with lamprey clotting factors. This is an evolutionary distance as great as it gets in vertebrates. It is known that protochrodates do not form fibrous clots, while all vertebrates form blood clots consisting of a fibrin matrix. This means the clotting cascade in its current form started to take shape when protochordates started to diverge from the lineage leading to vertebrates. This event coincides with the invention of red blood cells. These cells represent a large investment so that they should be retained in the body. In other words, clotting became important as blood "evolved into something more than another flavor of seawater" [31]. Through clotting it can be avoided that small injuries become fatal while a barrier is built for bacteria to enter the body.

In its most primitive form, the clotting cascade probably involved only six proteins. Four of those (tissue factor, factor X, (pro)thrombin, fibrin(ogen)) are also the key players in the modern clotting cascade. In different steps, this primitive scheme was embellished with more regulatory mechanism, and major changes can be attributed to four stages. These stages occurred at the splitting of the jawless from the jawed vertebrates, the

splitting of bony fish and tetrapods, the splitting of amphibians from the rest of the tetrapods and the splitting of marsupial from placental mammals. The multiplicity of the different protein domains indicate that the evolution of the clotting cascade builds on multiple gene duplication events. In such events, evolution can act on one copy of the gene while keeping a functioning mechanism in place [3, 31].

Jawless vertebrates have the simplest clotting system of vertebrates today [69]. But even this system has evolved over millennia from the simple six protein scheme mentioned above and now contains multiple clotting factors and mechanisms to localize the clot. The original clotting system has been lost to the past. While they have no intrinsic pathway, jawless vertebrates have a modified extrinsic pathway in which some clotting factors are missing while functional copies of others are present.

A large scale gene duplication, maybe even a whole genome duplication after the split of jawless vertebrates from the rest of the vertebrate tree paved the way for the more complex clotting cascades in higher vertebrates [69]. The first addition in this process was the development of the intrinsic pathway in jawed vertebrates [70]. The next step was the addition of the contact pathway to the clotting system of tetrapods [71]. This system is missing in fish which have modified extrinsic and intrinsic pathways. The addition of the contact pathway in tetrapods combined the clotting and inflammation systems. Interestingly the contact system has been lost in birds. After the divergence of reptiles from amphibians, a new regulatory mechanism was introduced into the clotting system. An antigen presenting protein in the cell surface was duplicated and re-purposed for activation of protein C. The most recent addition in the clotting system is the occurrence of apolipoprotein(a) in primates [31].

2.2.2 *Evolution of Fibrinogen*

The widespread occurrence of fibrinogen related domains suggests that the protein was present long before blood clotting was invented [72]. But once tied to the clotting system the evolution of fibrinogen is of course coupled to the evolution of the whole clotting system.

The most distant true fibrinogen known occurs in sea squirts [31]. In this organism, a likely dimeric form is found, complete with disulfide rings at the ends of the coiled-coil regions. This fibrinogen has three globular domains at the end of the coiled-coil region. An indication that the present version of the αC domain in vertebrate fibrinogen results from a runaway duplication of certain sequences of the FGA gene which destroyed the globular arrangement of the domain and increased its size. It is likely that only this modification of the αC domain allowed the evolution of fibrin fibers as they exist today.

The three chains in the sea squirt fibrinogen do not correspond to any of the chains in vertebrate fibrinogen. Instead, they are descended from the precursor of the α chain but have evolved independently over the last half billion years.

Interestingly, the sea squirt fibrinogen already has pockets in all three globular domains that are deep and structured enough for peptide binding. Yet no knobs are present in the central region so that association into fibers can not occur. The function of the pockets is presumably to interact with knobs present on the surface of cells. In this way, the fibrinogen molecule can connect to cells and start coagulation. Such a mechanism has been observed for the protein tachylectin 5A that is part of the immune response of the japanese horseshoe crab and has a γC like structure [73]. This protein clots foreign cells as part of the immune response. The clotting system in this case is not related to fibrin. Instead, it is derived from nerve growth factors. To effect the aggregation of cells, tachylectin 5A binds to carbohydrate groups on the cell surface. Such binding pockets for carbohydrates are indeed a common feature of fibrinogen related domains.

The occurrence of sea squirt fibrinogen as well as an analysis of mutations in fibrinogen suggest that the origin of fibrinogen is a single peptide chain that assembled along its coiled coil region. Later the different chains evolved through gene duplication events. The last event for the splitting of the β- and γ-chain is estimated to have occurred around 600 million years ago, close to the divergence of vertebrates and invertebrates and long before the last common ancestor of lampreys and humans lived [74]. Such analyses of mutations are complicated by the fact that the different parts of fibrinogen change at different rates. The coiled-coil region can undergo mutations frequently as the only constraint is that the association of this region is not disturbed too much. On the other hand, single mutations in the binding pockets can prevent the binding of the peptide ligand and severely impede fibrinogen function. Similarly, the disulfide rings are highly conserved.

As most other genes, the fibrinogen genes have evolved differently in different lineages of vertebrates. In bony fish, the αC domain underwent another major deletion event making it short. In contrast to this, the αC domain is larger in jawless vertebrates as it underwent a runaway gene duplication event. The jawless vertebrate also have a second gene for the alternative variant of the α chain that is produced by a different splicing of only one gene in other vertebrates. The genetic splicing variant giving rise to the alternative γ chain came into being after amphibians split from reptiles but only in mammals has this variant evolved a function.

2.3 FIBRINOGEN ADSORPTION

The physiological relevance of fibrinogen and its link to the contact pathway show the importance of protein surface interactions. In general, adsorption of proteins to surfaces is an important problem in medical and industrial applications [75, 76]. The contact pathway was just mentioned as a physiological example. Others are the adsorption of proteins at liquid-air interfaces that is important for foam formation in industry [77] or the adsorption properties of proteins for surface modification [78]. The focus of this work is on the adsorption from a water phase onto a solid surface.

In general, proteins adsorb at first reversibly on a surface so that they can easily change their orientation and maximize favorable interactions with the surface. After the optimal orientation is found, the interactions with the surface start to deform the protein at which point the adsorption becomes less reversible [79]. The denaturation of the protein also changes the protein dynamics which leads to a favorable entropic contribution to the free energy in the adsorbed state [80].

2.3.1 *Fibrinogens Relevance for Biomedical Materials*

The physiological importance of fibrinogen makes it a very relevant protein for medical applications. The study of the clotting cascade has for example lead to the application of tissue plasminogen activator as intervention to lyse clots in stroke patients [31].

The clotting ability of fibrinogen has also been put to use to stop bleeding during surgery. This treatment is so common that obtaining fibrinogen from the plasma of donated blood can be a problem and genetic engineering approaches are tested to produce it in large quantities in dairy cows [81]. In the design of new materials for wound dressings, patches of electro-spun fibrinogen have been used [82] and the incorporation of certain fibrinogen fragments in polymer patches was shown to increase wound healing by recruiting growth factors [83]. The interaction of fibrin with various growth factors has also been put to use in designed graft materials that promote the growth of osetoplasts [84] or bladder cells [85]. The unique mechanical properties of fibrin have likewise been used in innovative materials for vascular grafts [86]. In contrast to this, materials have been developed that are resistant to fibrinogen adsorption to prevent bacterial adhesion to, or clotting at, graft materials [87–89]. With all this, it is not surprising that fibrinogen adsorption can be used as a benchmark for the characterization of biomedical materials [90].

For internal applications of materials such as implants and stents fibrinogen adsorption is especially important. Not only is it the first protein to adsorb on the foreign material [91], it also adsorbs platelets well [92] and influences the behavior of

platelets and leukocytes that come in contact with the fibrin(ogen) matrix that is formed at the material interface [46, 93]. Such interaction can lead to the formation of a clot or an inflammatory response [94–96]. In the case of blood clot formation, thrombosis is the consequence, but worse complications arise if the clot becomes detached from the surface. For this reason materials are being developed that control fibrinogen adhesion and its interaction with platelets to improve haemocompatibility [97, 98]. Another effect that occurs when blood comes into contact with materials is the so called Vroman effect [99, 100]. This describes the fact that fibrinogen is replaced by other blood proteins on the materials surface after some time.

Nowadays the interaction of blood proteins with extended surfaces occurring in implants is not the only effect of consideration. In nanomedicine, i.e. the application of nanoparticles for therapeutic purposes such as drug delivery, the effect of the protein corona of a nanoparticle has to be considered. This protein corona can affect the final location in the body where particles end up as well as the internalization of particles by cells [101–103]. Additionally, it has to be avoided that nanoparticles coagulate, e.g. due to a fibrinogen molecule connecting two particles [104].

2.3.2 *Experimental Studies*

Due to its physiological and medical relevance, the adsorption of fibrinogen and its behavior in solution have been studied extensively during the last decades. In the following, the experimental literature is reviewed. Because bovine fibrinogen is more readily available, some of the studies use bovine instead of human fibrinogen. Since the fibrinogen sequence is evolutionarily highly conserved, the behavior is assumed to be similar in both cases and no distinction is made.

Detailed investigations of fibrinogen in solution are complicated by its size. These challenges are illustrated by the study of fibrinogen flexibility using fluorescent labels [105]. From the rotational relaxation of the fluorescence it was found that the data did not fit a stiff dumbbell model for fibrinogen. Instead, some flexibility in the connecting region between the two D-domains has to be assumed. Exactly which regions lead to this flexibility could not be observed. However, a later study showed that this interpretation is not conclusive and under different assumptions the data are compatible with a rigid fibrinogen molecule [106]. Recently the hydrodynamic characteristics of fibrinogen have been determined. The hydrodynamic radius for example was determined to be 12nm at acidic pH levels [107]. This number does not fit what one would expect from a stiff dumbbell. It was found that a partial detachment of the αC-regions from the fibrinogen core can explain the results [27]. However, the calcu-

lation of the hydrodynamic radius relied on the assumption of a stiff core in this case.

Due to these difficulties, most studies of fibrinogen have been performed at surfaces where imaging techniques like atomic force microscopy (AFM) or electron microscopy can be used. Such studies have the added benefit of shedding light on adsorption mechanisms that are relevant for biomedical materials. One interesting effect that is found in such experiments is that *in vivo* fibrin clots do not support strong cell adhesion whereas *in vitro* fibrin deposited at surfaces strongly adsorbs cells. The explanation for this effect is routed in the way cells regulate their behavior. If a cell attaches to a surface, it exerts forces on the substrate. Depending on the rigidity of the substrate the behavior of the cell is modified (mechanosensing). A plausible difference between *in vivo* and *in vitro* results is then the stiffness of the fibrin(ogen) matrix to which the cells attach. Indeed, it was found by atomic force spectroscopy that a monolayer of fibrinogen adsorbed to a surface is very adhesive while multilayers are successively less adhesive [108]. The adhesiveness drops sharply when the second fibrinogen layer is formed and increases again after multilayer fibrinogen matrices are cross-linked. This behavior can be understood in terms of the peeling off of fibrinogen layers. The first layer is tightly adsorbed to the surface, very likely accompanied by deformation that expose cryptic binding sites and strengthen adhesion. The second layer is only attached weakly to the first fibrinogen layer and its molecules deform considerably less. Hence a cell adsorbing to the second (or higher) fibrinogen layer that is not cross-linked to lower layers can not get a strong grip on the surface and easily detaches. This mechanism is speculated to regulate the growth of blood clots *in vivo*. If the clot is still growing, a substantial amount of fibrinogen is present that has not yet been activated by thrombin. If this forms a layer on growing fibrin fibers, it forms a protection against premature cell adhesion that would stop the fiber growth [46, 93]. Understanding this effect is also important to understand the dislocation and growth of fibrin networks under flow conditions. In a similar way, plasmin degradation of fibrin changes the mechanical properties of the substrate to which cells attach and can produce an anti-adhesive effect [109].

The question then arises which surfaces lead to the strongest adsorption of fibrinogen and what molecular characteristics of the surface influence the adsorption. One surface characteristic that has been explored is the hydrophobicity of the surface. It has been found that the adsorption is stronger at more hydrophobic surfaces [110–112]. In this case, fibrinogen also adsorbs faster and is less mobile on the surface. Another effect that influences adsorption is the electrostatic interaction between fibrinogen and charged surfaces like e.g, mica or latex [107, 113–120]. It is also known that the surface topology impacts adsorption. A results of this is the negative correlation between substrate crystallinity

and adsorption [121]. Other examples are the increased fibrin formation on rougher surfaces [122], the sensitivity of fibrinogen adsorption to nano-scale ripples (ripple wavelength 40nm–150nm) on the substrate [123], and the dependence of the adsorption on the curvature of nanoparticles [104]. In the case of ripple surfaces, the varying size of the ripples changes the curvature of the surface which in turn influences its hydrophobicity [123]. It also changes geometrical aspects: If the ripples are too close to each other fibrinogen will not be able to fit into the grooves and can form contacts only at the peaks, thus reducing the possible interaction energy. Curvature effects were also investigated using poly(acrylic acid) covered gold nanoparticles in the size range from 7nm to 20nm [104]. For particles of this size the curvature is significant even on the length scale of a single fibrinogen molecule . On the larger nanoparticles fibrinogen was more tightly bound while it was able to connect two nano particles for smaller diameters. The subtle effects the surface geometry and chemistry can have, make it important to study fibrinogen adsorption on materials that are actually used in medical interventions, not just on idealized model surfaces [124].

Once fibrinogen is adsorbed, the interaction with the surface can lead to denaturation of its tertiary and secondary structure elements. As such changes take time, the maximum adsorption strength is not reached immediately upon contact with the surface. An example of this was found on silica, where the maximum adsorption strength of fibrinogen is only reached after 50ms–200ms. Because the D- and E- region have a globular structure, their deformation can significantly increase the contact area with the surface by spreading and denaturation. It is thus not surprising that these regions are the major contributors to the adsorption energy [125]. On hydrophobic surfaces this spreading is more pronounced than on hydrophilic ones [126]. In the extreme cases, it has been observed that the height of the globular regions of fibrinogen slightly increases on mica and noticeably decreases on graphite [127]. The height increase on mica could indicate structural rearrangements, but the effect is small enough that it can not be excluded that the effect is due to measurement artifacts. The details of the structural rearrangements depend on the surface chemistry and results have been reported that show an increase in α-helices upon adsorption at amine terminated self assembled monolayers [117], while adsorption at other surfaces shows a decrease in α-helices and β-sheets [79, 128, 129].

As the structure of fibrinogen at a surface changes during the adsorption process, some binding sites for other proteins might be deformed beyond recognizability, while others become exposed in a similar way as they would in the fibrin fiber. It has thus been asked what functionalities fibrinogen retains after adsorption. For biomedical application the accessibility of integrin binding sites is important. With the help of a functionalized AFM tip, the $\alpha_{II}\beta_3$ integrin binding site located in the γ-chain of

fibrinogen has been probed in the adsorbed state on a mica surface [130]. It was found that the measured activity peaked after 45min at which point platelet adhesion also peaked. At longer times the activity decreased again. Similarly, it was found that this binding sites becomes accessible and stays active upon fibrinogen adsorption at hydrophilic and hydrophobic silicone surfaces [129]. In other studies, it was found that the P2 binding site of fibrinogen becomes exposed upon adsorption to poly(acrylic acid) covered gold nanoparticles [104, 131]. The adsorption of fibrinogen can furthermore influence its ability to form fibrin if thrombin is added. It was shown that somewhat surprisingly fibrin formation is not possible on mica while it is possible on graphite [132], despite the latter restricting fibrinogen mobility more [127]. The interaction of fibrinogen with thrombin is also affected on the polystyrene surfaces where the release of FpB is happening faster than in solution. At high surface concentration of fibrinogen, FpB is released even faster than FpA [133]. This effect might be due to the formation of aggregates of multiple fibrinogen molecules that disrupt the binding site responsible for FpA cleavage and form the site promoting FpB cleavage.

The difference in the accessibility of the fibrinogen binding pockets after adsorption to gold nanoparticles led to a model of end-on adsorption of fibrinogen at high concentrations in solution based on plasmon resonance measurements [134]. But without a direct visualization this adsorption mode remained contested and further research showed that fibrinogen adsorbs side on at all concentrations on mica [46].

Imaging Studies of adsorbed Fibrinogen

To understand the adsorption mechanism, knowing the structure of adsorbed fibrinogen layers can give valuable hints. Such structures can be derived from AFM, scanning tunneling microscopy, and electron microscopy. Attempts have been made to image fibrinogen using scanning tunneling microscopy, but not always has a single molecule resolution been achieved [135]. If single molecules could be visualized, they showed a trinodular structure [136]. Better results are obtained using electron microscopy. In fact, the earliest visualization of the fibrinogen molecule was done using electron microscopy of adsorbed molecules [18]. Even in those early pictures, the characteristic trinodular structure of fibrinogen is visible together with deviation from straight conformations. Later experiments confirmed this and showed deformed molecules more clearly [137]. Electron microscopy was also the tool of choice to reveal the structure of the protofibril [63]. The images in this study show aggregates of different sizes, ranging from three fibrinogen molecules to extended protofibrils. In this sequence, it became clear that single fibrinogen molecules often deviate from a linear structure but straighten out upon incorporation into an extended fiber. Recently AFM images have revealed another step in the fibrin fiber

formation. The two stranded protofibril is not always the smallest unit identified at surfaces. Rather a linear arrangement of fibrin(ogen) molecules, forming D-D-contacts, can be identified. This structures has been named a profibril [124, 138, 139]. It has been suggested that this association is driven by interactions of the αC-regions that can become decoupled from the E-region at surfaces [64].

By far the most prevalent method for the visualization of fibrinogen at surfaces is AFM imaging. Atomic force microscopy has the advantage that the surface can be imaged without further treatment. Today it is even possible to obtain high resolution pictures by performing AFM experiments in an aqueous environment. This technique has thus the potential to image fibrinogen adsorption at conditions that are closer to the physiological environment. While an AFM can nowadays take tens of pictures per second, this is still not fast enough to image the different stages of protein adsorption [140], and they have to be inferred from the final structures.

The surfaces most often used in AFM imaging of fibrinogen are graphite and mica. On both of them the trinodular structure of fibrinogen was observed [17, 127, 132]. This structure has also been seen on a variety of other surfaces [124, 126, 129, 141]. As the observed structures are not always straight, a bending angle was introduced to characterize the structures (see Fig 12). The angle is defined as the the angle between the three dots belonging to the same molecule. In some cases, the trinodular structure was not seen. Instead, a single blob was observed for a fibrinogen molecule [129, 142]. Only three studies have looked at the distribution of the bending angle. A study on graphite, made hydrophilic by glow discharge, found a single maximum at 180° [143]. A study on mica found a dominant peak at 157° and a smaller peak at 106° [17]. Recently, a third study was publsihed for fibrinogen adsorption on a graphite surface covered with carbohydrates [64] that yielded a similar distribution to the experiment on mica. The bending angle distribution on mica sheds some light on the origin of the blob structures observed in other experiments. The distribution extends to bending angles as low as 60° which corresponds to rather narrow triangular structures. Lower resolution studies would blur the three globular domains together so that only a single blob would be visible [17]. Earlier interpretation of the blob structures suggested that the αC-regions fold under the fibrinogen core during adsorption so that the adsorbed structure appears more compact [123, 124, 142]. In this model, the blob structures should dominate on hydrophilic surfaces such as mica because the αC-regions are themselves hydrophilic and trinodular structure should dominate on hydrophobic surfaces like graphite. This prediction is not in accordance with most imaging studies.

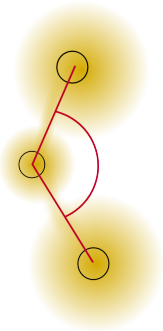

Figure 12: Representation of a typical AFM image of fibrinogen and the bending angle measured in these experiments

The high resolution of current AFM images also allowed the visualization of the αC-region in some cases and demonstrates

that they are not folded under the fibrinogen core during the adsorption on mica [17, 64]. The visualization of these domains has been reported previously, although, in this case the coiled-coil region have not been resolved so that it remains questionable if the αC-regions were seen or if an aggregate of two fibrinogen molecules was imaged [139]. High resolution images of protofibrils and fibrin fibers, in their initial stages of formation, demonstrate that the αC-regions are on the outside of the fiber [64]. This allows the αC-regions to drive lateral aggregation. That the αC-regions contribute significantly to the adsorption of fibrinogen has recently been demonstrated by comparing the adsorption of normal fibrinogen with fibrinogen without its αC-regions [144, 145]. The αC-regions contribute more to adsorption on hydrophilic surfaces than on hydrophobic surfaces. This is reminiscent of the earlier theory except that the placement of the αC-regions is not assumed to be between the fibrinogen core and the surface. Thus, the αC-regions do not give rise to more compact structures in AFM images while still providing a major contribution to adsorption on hydrophilic surfaces.

A prediction of this model is that the αC-regions would be attached to hydrophilic surfaces while they would be free on hydrophobic surfaces, assuming deformation in the E-domain release the electrostatic tethering. This means that via mechanisms discussed in section 2.1.3 fibers could assemble on the hydrophobic surfaces but not on hydrophilic ones. This is indeed what is observed in experiments [144, 145]. The fibers that are formed in this way are less stiff than actual fibrin fibers but have a similar structure. Addition of thrombin hardens the fibers and makes them denser. This process then gives rise to the characteristic band pattern of fibrin that is long known [17, 18, 60]. High resolution AFM imaging has also shown that the band pattern of adsorbed fibrin fibers becomes less pronounced as the fibers grow, indicating a less ordered arrangement of fibrin monomers in the outer layers of a fibrin fiber.

The variety of structures observed on surfaces and in solution experiments indicate that fibrinogen is very flexible in solution. While rough features of fibrinogen adsorption can be extracted from the described experiments, the size of the investigated system is often prohibits the identification of detailed mechanisms.

3

BIOMOLECULAR SIMULATIONS

In the last chapter, it became apparent that biological systems, such as the fibrin fiber, are so complex that it is not straightforward to perform experiments on them and interpret the results. This is due to a variety of phenomena that occur (nearly) simultaneously or so close together in space that they can not be distinguished in experiments. As a tool for understanding such processes, biomolecular simulations come into play that allow a high resolution picture of the process. The price one has to pay for such a detailed picture is the computational effort necessary to simulate sufficient system sizes and time scales to make meaningful inferences as to what is happening in experiment.

In this chapter, an overview over biomolecular simulation methods relevant to the topic of this thesis is given. Additionally, two major simulation techniques are introduced. After this, the CHARMM force field, developed for simulations of biomolecules in atomistic detail, is described before the multiscale problem and coarse graining approaches are addressed. The chapter closes with an overview of previous computational studies of protein adsorption in general and fibrin(ogen) in particular.

3.1 MONTE CARLO AND MOLECULAR DYNAMICS SIMULATIONS

The study of biomolecules using computational techniques started in the 1960s but has only seen a sharp rise in popularity during the last 15 years as its contributions did not only increase in number but also significance [146]. Leading the charge in these computational studies are the Molecular Dynamics (MD) and Monte Carlo (MC) techniques described in the following.

In this thesis only systems described by a time independent Hamiltonian $H(\{\mathbf{r}_i\}_i, \{\mathbf{p}_i\}_i)$ are investigated and the following considerations are limited to such systems. The description by a Hamiltonian implies that the systems evolves according to the Hamiltonian equations of motions and the phase space is symplectic. This means the dynamics are time reversible and the energy is conserved. Furthermore the phase space volume occupied by the system is conserved according to the Liouville theorem.

The quantities of interest for a thermodynamic study of such systems are averages of the form

$$\langle A \rangle = \frac{\int d\Omega \, A e^{-H/k_B T}}{\int d\Omega \, e^{-H/k_B T}}, \tag{3.1}$$

where $d\Omega$ is a volume element in phase space and k_B is the Boltzmann constant.

3.1.1 *Monte Carlo Simulations*

For a system of N interacting particles $d\Omega$ is given by $dr^{3N} dp^{3N}$, which highlights that the integral in eq. (3.1) is over a high dimensional space. In such a high dimensional space, most conventional numerical integration techniques are not efficient. An alternative is provided by Monte Carlo integration. In this method, points are randomly chosen from a closed region of the high dimensional space and the integrand is evaluated at these points. This allows an estimation of the integral.

The probability density

$$\rho(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N) = \frac{\exp\left[-\frac{H(\{\mathbf{r}_i\}_i, \{\mathbf{p}_i\}_i)}{k_B T}\right]}{\int d\Omega \, \exp\left[-\frac{H(\{\mathbf{r}_i\}_i, \{\mathbf{p}_i\}_i)}{k_B T}\right]} \tag{3.2}$$

becomes increasingly narrow as N increases. As a result it becomes more and more unlikely that a random phase space point would be in a region where $\rho(\{\mathbf{r}_i\}_i, \{\mathbf{p}_i\}_i)$ is non-vanishing [147]. A straightforward MC method would thus not be much more efficient than other numerical methods because myriad points would be necessary to obtain an accurate estimation of the integral. It is possible to improve this by not distributing the points uniformly in the given volume but focus them on regions where the integrand is non-vanishing. Such an approach is called importance sampling. For expressions of the type of eq. (3.1) this can be achieved by respecting the conidtion

$$\mathcal{N}_A P_{A \to B} = \mathcal{N}_B P_{B \to A} \tag{3.3}$$

for changes of the system conformation in subsequent MC steps. In eq. (3.3) $\mathcal{N}_i$ is the population of state i and $P_{i \to j}$ is the probability to go from state i to state j. This is the so-called detailed balance condition. In this way a Markov process is set up that samples phase space points according to the proability density in eq. (3.2). This is the Metropolis Monte Carlo algorithm and the basis for most MC simulations in statistical physics. An appropriate Markov process can be created by randomly changing the system conformation and accepting changes with a probability

$$P_{B \to E} = \begin{cases} \exp\left[-\frac{H_E - H_B}{k_B T}\right] & H_E > H_B \\ 1 & H_E < H_B \end{cases}. \tag{3.4}$$

The way in which new conformations are created depends on the system at hand and different update moves prevail for different systems [148].

3.1.2 *Molecular Dynamics Simulations*

Another way to evaluate thermodynamic averages is provided by the ergodic theorem. This theorem states that for a system in which Liouvilles theorem holds and trajectories reach almost all points (i.e. up to a subspace of measure zero) in phase space for almost all initial conditions the time and ensemble average are identical.

Thus, if the system is ergodic, the ensemble average in eq. (3.1) can be replaced by a time average of the form

$$\langle A \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, A(\mathbf{r}_i(t), \mathbf{p}_i(t))\,, \tag{3.5}$$

that can be calculated from a long trajectory of the system. This approach is the Molecular Dynamics method. The workflow for MD simulations is similar to experimental studies in that a system is first prepared and equilibrated before the measurement of quantities of interest is started. To determine the evolution of the system, the equations of motion are solved numerically. Because the position and momenta are the only phase space variables in this description, all measured variables have to be expressed as functions of these variables. Molecular dynamics simulations allow it to access dynamical quantities like diffusion coefficients or vibrational spectra.

The equations of motion are usually integrated using a Verlet integrator:

$$\mathbf{r}(t+\Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{F}(t)}{m}\Delta t^2 \tag{3.6a}$$

$$\boldsymbol{v}(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \tag{3.6b}$$

with a timestep Δt. This algorithm has the advantage that it can be derived on physical grounds as a discretization of the equations of motion [148]. It is time reversible and symplectic. thus the phase space volume is conserved even for finite time steps Δt. While the energy fluctuates on short time scales, the Verlet algorithm also ensures its long term stability, i.e. no drift in energy occurs on long time scales.

Because the simulations are supposed to be carried out in a canonical ensemble, an additional mechanism has to be added to introduce thermal fluctuations. The mechanism by which this is done is called a thermostat. Furthermore, a barostat has to be introduced for volume fluctuations. Examples for both are described later on in section 4.1.3. It has recently been shown that the modification of the dynamics by thermostats and barostats

can influence the correctsampling of phase space by the standard Verlet alogrithm [149, 150].

Because simulations can nowadays only capture microscopic systems, many assumptions of a large system in thermodynamics are not exactly justified. To alleviate the restriction to small system sizes, periodic boundary conditions are used and box sizes are chosen larger than the characteristic length scale of the phenomenon under investigation. In this way, a bulk system is approximated but an artificial periodicity is imposed. This can introduce artifacts in the simulation stemming from the limited box size, so-called finite size effects.

3.2 THE CHARMM FORCE FIELD

In MC as well as MD simulations, it is necessary to evaluate the energy (or its gradient) of a particle in a larger system. In principle, these energies arise from the quantum mechanical interaction of the electron clouds surrounding different atoms. While it is possible to simulate small systems on a quantum mechanical level (so called *ab-initio* methods), the calculation become too costly for larger system and some empirical function has to be used for the calculation of the energy and forces. Because these functions give rise to the forces in MD simulations, they are called force fields.

3.2.1 *Atomistic Force Fields*

In general, it is possible to define forces acting on many different subunits that do not have to be single atoms. This is the case, e.g. in coarse grained simulation discussed later. However, most commonly the term is used to refer to potential energy functions describing the interactions of (small groups of) atoms. In united atom force fields, non-polar hydrogen atoms are grouped together with the heavy atom to which they are bonded (usually carbon). In contrast to that, atomistic force fields represent each atom as an interacting point particle. The contributions to the potential energy are split into several different contributions representing interactions due to chemical bonds (bond distance, angle and dihedral) and non-bonded interaction such as electrostatics and van-der-Waals interactions representing the finite extend of the electron cloud. The bonded interactions are modeled as harmonic spring potentials and thus bonds can not break or form in standard atomistic simulations. Because bond formation can be an important step in certain processes, reactive force fields are being developed that use different descriptions of chemical bonds [151].

As each combination of chemical elements has their own bonding distance, angle, partial charge, etc. depending on the compound atomistic forcefields consist of myriad parameters for each relevant combination. These parameters are determined by

fitting results of computer simulations to structural data like X-ray crystal structures or NMR data, thermodynamic data like heat of vaporization or sublimation, and/or results of quantum mechanical calculations on small compounds (mostly for partial charges and dipole moments). However, this introduces a bias in the force field towards the properties used in the fit and it is a matter of active research how good the force fields are in reproducing novel situations. It has been estimated that an error in the interaction energy of less than 0.5 kcal/mol per residue is necessary to make reliable predictions [152]. One prominent example of biases toward the parametrization set is the dependence of partial charges of atoms on the environment. In current force fields, the partial charges are fixed and can not react to changes in the environment. The basic problem with fixed partial charges is their limited transferability so that improved determinations of the parameters for one case might not make a difference in another [153]. Because this behavior is unrealistic if polarizable substances like metal surface interact with charged groups, an effort is made to develop polarizable force fields [154]. A consequence of fixed partial charges is that force fields describe non-polar molecules like hydrocarbons often better than polarizable molecules [152].

The problem with assessing the quality of a force field is that a comparison with experiments is not always straightforward as conflicting results can be a consequence of insufficient sampling as well as wrong parameters. In the same way, force fields that seem to have a good performance on short time scales might show problems on longer timescales or in bigger systems when new conformational states are sampled. Furthermore, comparisons between different force fields can be difficult as the results will depend on the test system [152, 155, 156]. Usually the qualitative features and significant fluctuations match in all current biomolecular force fields but a quantitative agreement is not achieved. Furthermore, a comparison between quantum mechanical calculations and force fields for small proteins shows that the more costly *ab initio* methods can do better only for small proteins with disordered regions although in principle they correctly reproduce effects like bond-breaking, hydrogen bonding and partial charges [157].

In the past, several force fields have been developed for use with different MD programs. The major force fields currently in use for biomolecular simulations are GROMOS [158], AMBER [159], OPLS [160], and CHARMM [161]. Because only CHARMM is of relevance for this work, it will be described in more detail.

3.2.2 *The CHARMM Force Field*

The "**Ch**emistry at **Har**vard using **M**olecular **M**echanics" force field (CHARMM in the following) was originally developed in the group of Martin Karplus in the 1980s for simulations with

the program of the same name. Since then, it was adapted by more programs and has been extended to different systems by other groups.

In CHARMM, each atom is represented explicitly and assigned an atom type that determines which parameters are going to be used for it in the calculation of the potential energy. The potential energy in CHARMM is given by the following expression [161]:

$$
\begin{aligned}
V \;=\; & \sum_{\text{bonds}} K_b \left(r - r_0\right)^2 + \sum_{\text{valence angles}} K_\theta \left(\theta - \theta_0\right)^2 \\
& + \sum_{\text{dihedral}} K_\phi \left(1 + \cos[n\phi - \delta]\right) + \sum_{\text{improper}} K_\chi \left(\chi - \chi_0\right)^2 \\
& + \sum_{\text{Urey-Bradley}} K_S \left(S - S_0\right)^2 + \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \\
& + \sum_{i,j} \sqrt{\epsilon_i \epsilon_j} \left[\left(\frac{\frac{1}{2}\left(R_i + R_j\right)}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{12} - 2 \left(\frac{\frac{1}{2}\left(R_i + R_j\right)}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{6} \right]
\end{aligned}
\tag{3.7}
$$

The terms describe the following interactions:

BOND TERM: A harmonic potential where the parameters K_b and r_0 are specific to the type of bond, i.e. the atom types of the two atoms participating in the bond.

ANGLE TERM: A harmonic potential where the parameters K_θ and θ_0 are specific to the type of bonds connecting three atoms, i.e. the atom types of the three atoms connected by two bonds.

DIHEDRAL TERM: A periodic potential describing the torsion around a chemical bond. The parameter n determines the amount of stable torsional configurations. All parameters depend on the atom types of the four atoms defining the dihedral angle ϕ.

IMPROPER DIHEDRAL TERM: A harmonic restraint for the dihedral angle defined by four atoms that are not covalently linked. This term is used to ensure planarity around certain chemical bonds such as sp^2 bonds. The parameters depend on the four atom types involved in the definition of χ.

UREY-BRADLEY TERM: A harmonic potential for the distance between atoms separated by two covalent bonds. This term is introduced to reproduce in-plane deformations and asymmetric bond stretching modes.

ELECTROSTATIC TERM: A Coulomb potential to describe the electrostatic interaction between two atoms based on their partial charges. The partial charges are not linked to the atom type and depend on the larger compound to which the atom belongs. Typically they are fixed for a given residue.

VAN DER WAALS TERM: A 6-12 Lennard-Jones potential describing the van der Waals interaction of the electron clouds of the atoms. The position and depth of the Lennard-Jones potential are determined as the average of the values for an individual atom type.

The parameters entering the CHARMM force field have been optimized to reproduce gas-phase geometries, vibrational spectra, torsional energy surfaces, dipole moments, experimental heats and free energies of vaporization, solvation and sublimation, molecular volumes, and crystal pressures and structures, as well as *ab initio* results of water complexes with amino acids [161]. The parameterization was done using the TIP3P water model in which each atom is represented by a single interacting point [162]. Despite this parametrization, the performance of CHARMM depends only weakly on the used water model [156].

Subsequently the importance of the back-bone dihedral angles was noticed and numerical corrections were introduced. These correction were determined from *ab-inito* calculation and are referred to as CMAP corrections [163]. They lead to a reasonable agreement between protein folding simulations and experiment [164].

Recently the corrections to the CMAP values and the side chain dihedral angles have been introduced [165] and markedly improve the results that can be obtained from MD simulation with the CHARMM force field [155]. That such improvement in secondary structure balance are necessary had been noted before [164]. Employing corrections for improvements of the CMAP term the CHARMM force field was used in landmark protein folding simulations [9]. These simulations reached the timescale of milliseconds and showed repeated folding of proteins into their experimental structures.

Originally developed for proteins CHARMM was soon extended to lipids and nucleic acids. Parameters for these three classes of biomolecules are the core of the CHARMM force field. As more complex systems are studied nowadays, the necessity arises to extend CHARMM to other materials. This has been done for small, drug-like molecules [166] and carbohydrates [167, 168].

For materials science the interaction with surfaces is of great importance and to this end force fields for different classes of surfaces have been introduced. These materials include silica surfaces [169], clays [170], and metals [171]. As a target for parametrization, experimental observations of surface effects such as water contact angles and adsorption energies of peptides are usually chosen, but *ab-initio* calculations can contribute information on the parameters as well [172]. The transferability of the force field parameters for surfaces is an even bigger issue than for solution simulations of proteins [173] and in some cases the force field is parameterized to reproduce the interfacial behavior

exclusively to the detriment of other materials properties of the interface that depend on bulk dynamics.

3.3 COARSE GRAINING AND THE MULTISCALE PROBLEM

The development of force fields for protein simulations is based on the fact that it was and still is not possible to simulate the large system sizes required to understand protein dynamics on a fully quantum mechanical level. The reason is that *ab-inito* methods need to solve the Schrödinger equation in order to evaluate the energy of a certain conformation and the forces acting on it. This is vastly more expensive than evaluating functions like the one in eq. (3.7) in classical simulations. On the other hand even in large biomolecules the breaking of a bond or small conformational changes due to adsorption of a photon might be the key event for large structural changes and are lost in atomistic simulations. Examples are the cleavage of glycosidic bonds by lysozyme [175], the photosensitivity of rhodopsin [176, 177], and, more generally photosynthesis and photovoltaics [178].

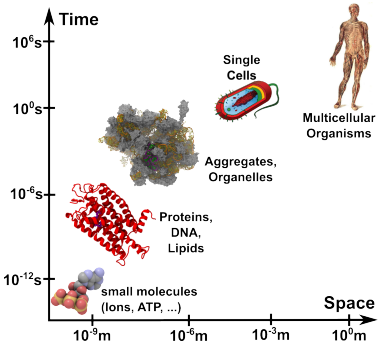

Figure 13: The multiple interconnected time and length scales in biological systems range from small molecules like ATP to whole organisms and from dynamics on the femtosecond scale (bond vibrations) to years. [174]

Even if classical simulations are performed, the timescales that are typically reached nowadays are limited to hundreds of nanoseconds and systems sizes on the order of 10^6 atoms which are still too small to understand many biological phenomena. An illustration of the multiple scales in biological systems is given in Fig. 13. One approach to reach longer time scales is to develop so called coarse grained (CG) models in which several atoms are grouped together in a coarse grained bead. This reduces the degrees of freedom and allows one to follow the large scale motions over longer periods of time. The reduction in degrees of freedom leads to a smoothing of the free energy surface that eliminates many metastable states. This allows simulations to explore the free energy surface faster, an important advantage considering that proteins have a very rugged free energy landscape with multiple, long-lived meta-stable states similar to glasses [12, 13]. The first coarse grained model for a protein was proposed in 1975 by Levitt and Warshel to understand protein folding [8].

The examples mentioned above demonstrate that the dynamics of proteins happen on multiple time and length scales where each level of resolution affects the next. This is a so called multiscale problem that occurs frequently in biophysics and soft matter[1] physics more generally. Simply stated the problem is the following: The large scale behavior of a system is to be investigated, but this behavior is linked to smaller scales that are too costly to resolve. Developing methods to bridge different scales has been identified as one of the key challenges for the modeling and simulation of biomolecules [146] and in 2013 the pioneers of multi-

1 In soft matter systems the relevant energy scale of the system is $k_B T$ so that thermal fluctuations can cause large scale changes. This is frequently the case for biomolecules at room temperature.

scale methods in biomolecular systems, Martin Karplus, Michael Levitt and Arieh Warshel have been awarded the Nobel Prize in Chemistry "for the development of multiscale models for complex chemical systems" [179]. The importance of multiscale modeling and simulation is not only expressed through the Nobel prize but also the wealth of reviews summarizing advances in the topic [146, 180–186]. Furthermore, the multiscale dynamics also requires analysis tools that combine views of large conformational changes with local fluctuations [187] and even if simulation become fast enough to study high resolution models of large systems, coarse graining methodologies might be essential in distilling the important effects out of a heap of data [182]. Here some of the methods for multiscale modeling of biomolecular systems are presented.

3.3.1 *Multiscale Modeling*

In the following, multiscale modeling is meant to represent approaches that try to achieve a quantitative agreement between different levels of description by a systematic derivation of a coarse grained model and its parameters from the study of the fine grained system. The methods for multiscale simulations can be grouped into sequential and concurrent approaches. In a sequential approach, information from one level is passed to the next and only systems of one resolution are simulated. In concurrent approaches, two or more levels of description are present in the same simulation. This can be the case for fixed particles (hybrid methods) or change depending on certain conditions (adaptive resolution methods) [183, 186]. In principle, the force field development described in section 3.2 is an example of sequential coarse graining in which the electron degrees of freedom are eliminated. The corresponding hybrid method connecting the same levels of description is QM/MM (quantum mechanic/ molecular mechanic) where some atoms in a system are treated quantum mechanically while others are treated on a classical level. This approach works best if the QM and MM regions are separated in such a way that no particles or bonds cross the QM/MM boundary [188].

When a coarse grained model is constructed, the first step is to define what constitutes a CG bead. This can be an atom in the case of atomistic force fields or a whole protein domain for highly coarse grained models. Different approaches exist as to define a mapping from a fine grained to a coarse grained model. At low coarse graining levels, chemical intuition might be enough to identify CG units as is done, e.g. in the MARTINI model where amino acids are represented by 1–5 beads [189, 190]. This is enough reduction to allow simulations of large protein aggregates like viral capsids [191] but not enough for other systems. The higher the coarse graining level the more desirable a systematic way of defining CG beads becomes to eliminate re-

searcher biases. This can be achieved to some degree with neural network approaches called self organized maps [180], identification of rigid domains [192] or essential dynamics coarse graining [193] (described in detail in section 7.1).

In adaptive resolution methods, particles can change their level of description and pass from higher to lower resolution models and vice versa. Because adaptive resolution models like the AdResS method [181, 194] are not used in the present thesis, they will not be discussed here. Nevertheless, these methods highlight some difficulties for consistent multiscale models because they couple the thermodynamic systems of both resolution levels. The most obvious of these challenges is the backmapping problem. Backmapping is the procedure that makes it possible to go from a low resolution model back to a high resolution model. Because each CG bead corresponds to an ensemble of high resolution structures, this backmapping is possible if the local relaxation is faster than larger scale structural motions [186]. The backmapping problem also occurs if the coarse grained simulation is used to speed up simulation in order to create new large scale conformations as starting points for further high resolution simulations. Such an approach has been used in the case of protein insertion into lipid membranes [195].

The condition that local relaxation should be faster than structural motions for the backmapping illustrates that there should be some separation of time scales for the relevant processes in the system. Indeed, if time scales clearly separate, it is possible to systematically derive evolution equations for the CG degrees of freedom from the Zwanzig or Green formalisms [196, 197]. If no separation of time scales exits, the Zwanzig formalism gives evolution equations that contain a memory kernel, which in practice can not be calculated. Unfortunately for internal motions of proteins no time scale separation seems to exist [198] so that methods with a less well established thermodynamic background have to be used and should be validated on a case by case basis. However, the memory kernel that according to Zwanzig exists for CG variables in proteins can be used as a measure for the quality of the CG model. The better the CG model is the closer this kernel should be to a delta function [199]. Another way to assess the quality of a CG model is the relative entropy suggested by Shell [200]. The relative entropy quantifies the overlap between the thermodynamic ensembles of the high and low resolution models and should be minimized in an optimal model.

A further problem arising from the lack of separation of time scales is the CG time itself. One of the main points of smoothing the free energy landscape with coarse graining is to speed up dynamics. Unfortunately this speed up is not always uniform and some processes are accelerated more than others so that the interpretation of CG time becomes problematic and it is safer to see the simulations as a sampling of phase space instead of coarse dynamics [182, 186].

Another important criterion for a CG model is consistency with the fine grained model. This is meant in the sense that the behavior of a CG bead should correspond to that of its constituent fine grained particles. This could be achieved by determining interactions so that structural characteristics like the pair correlation function are reproduced. This means the pair correlation function of the CG beads is the same as the one for the centers of mass of its fine grained building blocks. This approach is based on Henderson's theorem [201] that states that there is a unique potential (up to a constant) for each pair distribution function. It is known that there can be problems in its application in practice [202]. A widespread method to calculate the potential that produces a given pair distribution function is the Iterative Boltzmann Inversion [203]. In the dilute regime, the pair correlation function and the potential are related via

$$V(r) = -k_B T \ln [g(r)] \ . \tag{3.8}$$

For denser systems this is not correct anymore as multi-particle interactions become important and the potential calculated in eq. (3.8) is called the potential of mean force. To recover many particle effects, this equation is iterated according to

$$V_{CG}^{n+1}(r) = V_{CG}^{n}(r) - k_B T \ln \left[\frac{g_{CG}(r)}{g_{FG}(r)} \right] \tag{3.9}$$

until the coarse and fine grained results agree. Alternatively the Inverse Monte Carlo Method [180] can be used to determine the potential.

A different approach to determine the CG interactions is force matching which was developed to parametrize atomistic force fields from *ab-initio* calculations [204–206]. As the name suggests, this approach varies the forces in the CG model until they match the fine grained model. This is formalized by a minimization of

$$\chi^2 \propto \sum_{l=1}^{L} \sum_{i=1}^{N} \left| \mathbf{F}_{il}^{FG} - \mathbf{F}_{il}^{CG} \right|^2 \ . \tag{3.10}$$

A drawback of this method is that the choice of functions as a basis set for $\mathbf{F}^{CG}$ determines the quality of the resulting interactions. Like the Boltzmann Inversion, force matching only considers pairwise forces so that an iterative procedure has to be employed to get a good fit [207]. It was also shown that interactions determined in this way will give rise to a CG model that has a phase space probability distribution that corresponds to that of the projected fine grained system [208]. Furthermore, it can be extended to a model where CG beads have internal states [209], a fact that becomes important in highly coarse grained ($\gg$1 amino acid per CG bead) models of biomolecules, where, e.g. small rearrangements of residues can change the affinity of binding pockets.

Both backmapping and thermodynamic consistency are essential to allow multiple transitions from one level of resolution to the next and use results from one to refine the other in an iterative process [182]. The drawback of such consistent models is the great computational effort necessary to parametrize them as $g_{FG}(r)$ or $F^{FG}(r)$ needs to be known and the corresponding CG quantities have to be calculated repeatedly. Even then, the same transferability problems as discussed for force fields arise for the CG interactions. The parametrized interactions might just reproduce the dynamics at a certain temperature, ion concentration and so on [186]. Depending on the questions asked it might be advantageous to construct less rigorous phenomenological models that can only qualitatively reproduce experiments [183]. Often the distinction between rigorous multiscale models and phenomenological models is far from sharp as restrictions on computer time necessitate short cuts in the CG model and force fields and phenomenological models are parametrized to reproduce simulation and experimental data well, even if no thermodynamic backing for the procedure is established. In the following, more phenomenological models for the simulations of proteins are reviewed.

3.3.2 *Phenomenological Approaches for Proteins*

In a coarse grained model, the smoothed free energy landscape means that a longer timestep can be used in MD simulations. The length of the timestep plays an important role in sampling a system over extended periods of time: A longer time step allows to sample long trajectories with less computational effort. The slower the fastest mode in a system is the bigger a timestep can be chosen. This is nowadays frequently used in biomolecular simulation where the length of covalent bonds involving hydrogens is fixed. This eliminates their fast vibrations and allows a doubling of the timestep from 1fs to 2fs. This approach has also been extended to fixing other bonds [210] thus coarsening the dynamics without coarsening the structure.

A natural level of resolution for coarse grained models of proteins is the amino acid level. At this level, each amino acid is represented as a single CG bead and over the years many such models have been developed [184, 185]. Elastic network models (ENM) have the simplest form of interaction between the CG beads. In the earliest concept, a bead was connected to all other beads within a certain cutoff using harmonic springs of the same spring constant. The spring constant was then adjusted to reproduce fluctuations observed in experiments (e.g. temperature factors in X-ray crystal structures). Because the elastic network is based on the crystal structure, this introduces a heavy bias towards this structure and only fluctuations around it can reliably be reproduced. Later these models were generalized to include multiple stable structures, plastic deformations and het-

erogeneous spring constants [211, 212]. Because all interactions are harmonic, the motion of the protein can easily be investigated for biologically relevant motions using normal mode analysis. The normal modes of an elastic network model can also be used as a way to further coarse grain the protein [213]. In this approach, the motion of the protein is approximated by a time dependent overlap of the dominant ENM modes and the time dependent coefficients are treated as the CG variables. A recent study confirmed the relevance of the dominant modes of anisotropic elastic network models on the milli- to microsecond time scale [214].

The Gō-models are another type of important residue based CG models for proteins and were developed for protein folding. The potential energy in these models is usually more complicated than in elastic network models and contains harmonic bond and angle terms in the backbone as well as Lennard-Jones interactions. Only amino acids closer than a certain cutoff in the crystal structure have an attractive part to the Lennard-Jones interaction in this model. All others interact with a purely repulsive potential. The interaction energy is fixed in such a way as to reproduce experimentally observed melting temperatures of the structures. In a similar model for disordered protein regions, the C_α atoms were connected by springs and interacted via a purely repulsive potential while only native contacts had a short ranged attractive interaction [215]. The parameters in the model were adjusted to reproduce experimental data on the radius of gyration of the protein.

The next step in the CG hierarchy would be to group multiple residues together. This relies on the identification of rigid domains in protein structures. One approach to this without the need for expensive atomistic simulations has been advanced by Thorpe and co-workers [192] who created rigid domains by placing constraints between residues connected by covalent or hydrogen bonds, salt bridges, and hydrophobic interactions. The finite dimension of the domains was reconstructed using Lennard-Jones interaction centers at the C_α position of each residue. The dynamics of the rigid domains was recovered by analyzing the underlying elastic network model. A similar approach has been proposed by Hummer and co-workers who treat domains as rigid structures and each C_α atom as an interacting bead. The beads interact electrostatically via a Debye-Hückel model and have an additional Lennard-Jones interaction. The Lennard-Jones interaction is parametrized to reproduce binding affinities. This model is useful to investigate protein-protein interactions [216] as well as structural investigations of proteins with large flexible parts [217].

As proteins are polymers, it is natural that concepts from polymer science have been used as CG models for proteins. One challenge is that the secondary structure of proteins is a more pronounced local structur than what occurs for many polymers.

This was addressed by Levine who introduced an internal field to a wormlike chain model of a polymer. This internal field was Ising-like in that it could only take two values and had nearest neighbor interactions [218]. The values of the field represented an α-helical secondary structure or a random coil and influenced the persistence length of a certain part of the polymer chain. Because α-helices are stiffer than coils, the persistence length was greater in these regions. With this model, it is possible to investigate the behavior of long helical segments or coiled-coil regions under external forces such as bending or stretching. Under force it is known that the helical conformation can be broken into a coil and in Levines model this becomes equivalent to the nucleation of a second phase in the Ising field. Once the nucleation happens, the curvature of the helical segment is no longer uniformly distributed over its length, rather a localized kink appears. Similar investigations show that the extension under an applied force can be divided into four regions: The straightening of the chain, extension of the helical region, breaking of the helical structure and extension of the coiled region [219, 220].

On an even higher level proteins have been described as single beads to study their translocation through nanopores [221, 222] or the environment has been simplified [223]. In this case, no matching to a microscopic model was done beyond conservation of the total charge.

3.4 PROTEIN AND PEPTIDE ADSORPTION SIMULATIONS

One example where a multiscale approach is necessary is the adsorption of proteins on surfaces. On the one hand quantum mechanical calculation are necesary to understand these phenomena. This is necessary to get a complete picture of the interactions of the protein with the surface and parametrize forcefields. On the other hand atomistic resolution is desireable to obtain larger conformation of the molecule at the surface, as well as to resolve the interaction between the solvation layers of the protein and the surface. Coarse grained techniques are necessary to address layer formation or the growth of nanoparticles. In section 2.3, the importance of protein adsorption at surfaces was discussed for the case of fibrinogen. Of course fibrinogen is not the only protein of relevance for adsorption phenomena, and many other proteins have been studied for their effects in biofouling of industrial materials or biomineralization processes. As a consequence, the adsorption of proteins at surfaces has attracted not only experimental but also computational scientists [173, 224]. Therefore, a more general overview of protein and peptide adsorption simulations is given before previous computational studies of fibrinogen are presented in the next section.

The sampling of sufficient conformations of a protein-surface system is even more difficult than in solution because a molecule can approach the interface in different orientations. If the orien-

tation is unfavorable, the protein can go back into solution and even if a favorable orientation approaches thermal fluctuation can lead to desorption events. Because it is not practical to simulate relevant system sizes long enough to observe approach, desorption, reorientation and a new approach in a single simulation, proteins are usually placed in different initial conformations close to the surface [225–230]. In this way, different orientation can be judged and preferred adsorption orientations can be identified. Placing the protein close to the surface ensures that no undue amount of simulation time is spent on the diffusive approach of the protein to the surface. Conformational changes in the surface lead to additional sampling problems. While in hard inorganic surfaces only local vibrations occur, soft polymer surfaces can show significant rearrangements due to thermal fluctuations. For such a soft interface it can already be a challenge in itself to find a starting structure for the surface [231, 232]. To reduce the sampling time and mimic the effect of a bulk material, the dynamics of (parts of) the surface is usually constrained. This is done by fixing the position of the atoms to their initial values [169, 228, 233–236]. No real bulk material is simulated because the periodic boundary conditions lead to a slab geometry for the surface, i.e. two interfaces are present in the simulation box. While a sufficiently large extension of the box perpendicular to the surface can ensure that the protein only interacts with one interface, it is not clear that long range interactions of the water layer below the solid slab are unimportant for adsorption [237]. Hence, the choice of the slab thickness has to be given some thought. In practice, this thickness is chosen to be in the range of 1.4nm–2.5nm [225–227, 230, 238] although smaller [233] and larger [228] values have been used.

While an explicit representation of the solvent is potentially very important for adsorption phenomena because both protein and surface are surrounded by a hydration layer [227–229], it is not always practical to use it. Especially in coarse grained models it is common to use implicit solvent environments but also atomistic studies have used it [234–236]. Other ways to speed up adsorption simulation include steered Molecular Dynamics where an external force is applied to the system, pushing the protein towards the surface [227] or MD runs interspersed with energy minimization procedures [233–236].

The surfaces investigated in simulations range from models with a pure Lennard-Jones potential to atomistic representations of medically relevant materials like Titanium. An overview over some examples of the different surfaces used for biomolecular simulations is given in Table 3. Most of the studies mentioned here used MD simulations. The exceptions are [229] and [239] where MC simulations were used. Contacts between the proteins and the surface were identified as protein atoms within a certain distance of surface atoms (usually 0.5nm [228, 230, 231]) although an energetic definition of a contact is also possible [244].

Surface	*Reference*
pure Lennard-Jones	*[229], [239]*
Graphite	*[227], [233], [234], [235], [236]*
Mica	*[225], [238]*
SiO_2	*[240], [241]*
GaAs	*[239]*
MgO	*[228]*
Polymers	*[232], [233]*
Lipids	*[242]*
Titanium	*[227]*
Gold	*[225], [226], [230], [243]*

Table 3: Examples of the different surfaces used in atomistic biomolecular simulations.

The identification of an optimal adsorption orientation can help to clarify adsorption mechanisms. As an example, it was shown that peptides show a weak epitaxy on metal surfaces where the side chains arrange so that they fit the epitaxial centers on the surface [226]. The different orientations of peptides on mica and graphite show that the hydrophobic graphite surface promotes a spreading of the peptide on the surface without the formation of aggregates while the structure of mica leads to dense upright assemblies. In these assemblies, the molecular packing leads to the formation of β-sheets and could be a precursor for amyloid formation [238].

Structural changes upon adsorption like the β-sheet formation just mentioned, are another topic for computational investigations. In general, it is found that structural changes are more pronounced on hydrophobic surfaces where it is favorable to bring the hydrophobic core of a protein in contact with the surface [231, 233]. Such changes might happen by rearrangements of the tertiary structure with only minor changes in the secondary structure [234–236]. As these changes take time, it can happen that no changes of the protein structure can be observed on the simulated time scale [227]. It has also been found that different secondary structure elements show different adsorption strength on the surface [239]. To characterize the adsorption strength, the number of contacts with the surface or the adsorption energy can be used and accurate methods to calculate this energy using simulations have been developed [225].

Furthermore, the effect of nano structures on proteins has been investigated. In the case of carbon nanotubes, the main structural feature is the curvature of the surface and it has been shown that an increase in curvature leads to weaker protein adsorption on carbon surfaces [231, 245]. For silica nanoparticles of different sizes the curvature is not the only thing that changes. Particles of different size also differ in the topology and degree of ionization of their surface. The effects of such changes for peptide adsorption have recently been investigated to show that silica nano particles of different size adsorb different peptides [241].

3.5 PREVIOUS COMPUTATIONAL STUDIES OF FIBRINOGEN

As was mentioned in section 2.3.2, the identification of molecular details is difficult in experiments on fibrinogen and fibrin. Therefore, theoretical studies of fibrinogen have been performed to supplement and illuminate experimental data. In the following, the literature on theoretical and computational descriptions of fibrinogen is reviewed.

On a very high structural level the effect of fibrin rupture on plasma flow has been investigated in a very simple model in which blood cells are connected by bonds that can break due to overstretching. A random bond breaking was also included

to mimick the action of plasminogen. Such studies can help to explain the size of dislocated clot fragments in the blood [246]. A step lower in the resolution are kinetic models that describe the formation of fibrin as a sequence of reactions with different rate constants [247]. The rate constant can at the moment only be obtained from experiment and a numerical study of the model is possible. Such a model shows that a simple three step process consisting of FpA cleavage, protofibril formation and fiber formation as discussed in section 2.1.3 is not sufficient to reproduce experimental results. Instead, an additional step has to be added that requires the protofibrils to have a certain length before they can aggregate to fibers. the problem with this approach is that the structure of the used reaction network is not clear *a priori* but needs to be adjusted to fit experimental observations. However, once a satisfactory network is found the effect of different rate constants can be investigated to predict fibrin formation under different physiological conditions. Similarly, the structure of fibrin networks has been modeled. These models show that the networks have the structure of a fractal network of randomly connected cylinders [248].

To predict the behavior of fibrin, it is important to understand its mechanical properties on smaller scales. These are obviously rooted in the mechanical properties of fibrinogen that are characterized by its deformation under force. Force extension curves of a single molecule can be recorded using AFM techniques. However, for a molecule as complex as fibrinogen it is almost impossible to interpret these measurements without simulations that assign certain features in the force extension curves to structural changes. An atomistic simulation of such a kind has been performed on fibrinogen and compared to experiments [249]. Three distinct steps in the forced unfolding process have been observed: In line with Levines model the protein straightens first under the applied force, then the coiled-coil regions extend and unfold, and finally the E-domain unfolds, i.e. the β-hairpin of protomer is pulled out of the loop formed by the second protomer. The simulations were performed as steered MD simulations, where the pulling force attached at the end of the coiled-coil region so that the D-region remained undisturbed. The results have been reproduced using a coarse grained model of fibrinogen [211]. This model was a dynamic elastic network model in which the only interactions were harmonic springs connecting the C_α atoms of the amino acids. These springs could be removed when the force exerted on them became to large. This represents bond breaking or unfolding and allows to study force extension behavior of proteins represented by an elastic network model.

In fibrin, the force will not attach at the end of the coiled-coil region. Instead, linear arrangements of fibrin monomers, cross-linked at the D-D interface will be pulled. Such arrangements are too large to study in atomistic detail and a coarse grained model

has been used to investigate them [250]. The coarse grained model consisted of a self organizing polymer model [251] in which each residue is replaced by a bead. The beads are connected with finite extensible nonlinear elastic (FENE) potentials along the backbone and Lennard-Jones-like potentials for longer range interactions. This model allowed slower pulling speeds and is in good agreement with experimental results. It was shown that the γC-domains unfold in a multistep process after the coiled-coil regions are sufficiently extended but not yet unfolded. Each unfolding step in the γC-domain leads to a relaxation in the coiled-coil region. The βC-domain is not affected, as it is not in the line of force transmission. The D-D interface was found to unfold easier than the γC-domain, so that in a first step the D-D interface separates before the unfolding process continues as it would for a single fibrinogen. A next step would be to include the D-E interactions in this model. The first step in this direction has been taken by investigating the knob-hole interactions [252]. Starting from the crystal structure with bound tetrapeptide knobs the D-region was simulated in atomistic detail for 10ns before the knobs were pulled out of the pockets using steered MD. The simulations studied the pocket dynamics upon extraction of the ligand as well as the Gibbs free energy profile associated with the process. In the simulations, different temperatures and pH levels (defined by the protonation state of the amino acids) were investigated. It was found that the interaction of the A-knob with the γ-hole was not stronger than the interaction of the B-knob with the β-hole. This is at odds with experiments and in a follow up study of the D-E-D complex indications were found that more than the first four peptides of the knobs are involved in binding. However, the simulation of the D-E-D complex could not be regarded as conclusive. Even without detailed information on the D-E force extension behavior, a model of the fibrin fiber could be build that explains its strain hardening [253]. This model consists of a fibrinogen core that has the force extension behavior measured in [249] and αC-regions that are modeled as wormlike chains. The fibrin fiber can then be modeled as a network of springs and the behavior of this network can be investigated numerically [253].

A drawback of simulations of the force extension behavior of proteins is that due to the limited time that can be simulated pulling speeds have to be used that are much greater than in experiments. In a typical AFM study of fibrinogen extension behavior, the pulling speed is in the range of 0.4 $\frac{\mu m}{s}$ to $1\frac{\mu m}{s}$, while in [249] the pulling speed was $2.5\frac{m}{s}$ and in [211] it was $1\frac{km}{s}$. Only in [250] a realistic speed of $1\frac{\mu m}{s}$ was reached.

As many experiments have studied the adsorption of fibrinogen, so have some theoretical approaches. The only atomistic study to date looked at the γC-domain of fibrinogen at different self assembled polymer layers of 1nm thickness[232]. The γC-domain was initially placed in such an orientation that the

contact area of the whole fibrinogen molecule would be maximized with a minimal distance of 0.7nm to the surface. It was found that the hydrogen bonding and hydrophobic interactions dominate the adsorption behavior and that water layers are retained between protein and surface. Furthermore, it was found that the domain stays rigid and starts rolling along the surface.

Since the simulation of layer formation is not possible in atomistic detail, coarse grained models have to be used. One approach that has been taken is too approximate fibrinogen as an ellipsoid to gain insights into the mechanisms underlying the Vroman effect (see section 2.3.1) [256]. Another approach is the random sequential adsorption (RSA) method that was applied by Adamczyck and co-workers to fibrinogen and other proteins at surfaces [254, 255, 257]. The RSA method is used to determine surface coverages by placing objects sequentially at random positions on a surface. If too many attempts at placing an object on the surface are unsuccessful, the procedure is stopped and it is assumed that the maximum coverage of the surface is reached. In this method, fibrinogen is treated as a sequence of rigidly connected beads of different size with the sole purpose of preserving its shape as good as possible (see Fig 14). It was assumed that fibrinogen will adsorb initially in a side on conformation and at higher surface density will change to end-on adsorption based on the experimental results in [134]. The RSA method can also be extended to take transport effects on the surface into account which is important as they can be more important than the geometric restrictions of the fixed rigid bodies. Using this model the surface coverage of fibrinogen and its adsorption isotherm can be calculated and compared to experiments. The interactions energies obtained by comparison with experiment suggest that fibrinogen should adsorb irreversibly in the side-on orientation while adsorption in the end-on orientation is reversible. The model has recently been extended to include the αC-regions and predicts a brush-like structure for the fibrinogen layer on latex nanoparticles [27, 118]. Another extension has been made to include partial ordering in the adsorbed fibrinogen layer [258].

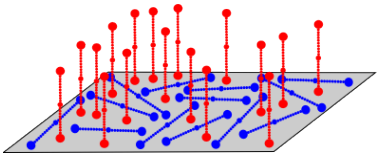

Figure 14: Representation of the fibrinogen model studied by Adamczyck using RSA [254, 255]. Fibrinogen is modeled as a string of rigidly connected hard spheres that can be placed on the surface in end-on or side-on orientation.

In conclusion, it can be said that computational studies are nowadays used more and more to explain results of experiments on biomolecules. Especially at surfaces improvements have been made in the development of force fields that can describe the protein surface interaction. Together with the increased development of systematic coarse graining strategies, techniques are at hand to understand and characterize the behavior of proteins in solution and at surfaces and use this knowledge to develop models that carry the desired chemical characteristics of higher resolution models. In the rest of the thesis, this strategy will be applied to fibrinogen.

Part II

ATOMISTIC SIMULATIONS

4

METHODS

To understand the dynamics of fibrinogen on an atomistic scale, be it in solution or at interfaces, Molecular Dynamics simulations are performed. The steps taken to perform such simulations are described in the first part of this chapter. The second part of the chapter is concerned with the analysis tools and methods that are needed to extract meaningful information from the collected MD data.

4.1 SIMULATION DETAILS

As described in section 3.1.2, Molecular Dynamics (MD) simulation sample the phase space of a system by evolving its dynamic for a sufficiently long time. Because this gives not only a hint of the important states of a system, but also the dynamical processes connecting these states, it is usually the tool of choice for the investigation of biomolecules and will be used in this thesis.

4.1.1 *Simulation Set Up and Preparation for Fibrinogen in Solution*

The starting structure

As a starting point for the atomistic MD simulations, the crystal structure of human fibrinogen [19] was used. The structure was retrieved from the PDB entry 3GHG. It contains two fibrinogen dimers of equal resolution from which the one with fully resolved carbohydrate groups attached to the Bβ-chain was chosen.

To identify potential artifacts resulting from the crystal structure it has been checked that no atoms were resolved with an occupancy smaller than one. While no occupancy smaller than one occurs for the fibrinogen core, several residues at the N- and C-termini are missing. The unresolved residues in the used structure are given in Table 4.

Chain	*Missing residues*
α	*1–26, 201–562*
β	*1–57, 459 –461*
γ	*1–13, 395 – 411*

Table 4: Unresolved residues in the fibrinogen crystal structure

Only 31% of the α-chain have been resolved. The missing residues correspond mostly to the αC-region. It is well known that this domain is a very flexible part of the protein and is mostly devoid of secondary structure elements [21]. The flexibility of the αC-region has hampered efforts to crystallize fibrinogen. Even today, it has not been possible to crystallize fibrinogen without removing its αC-region, let alone resolve all of the domain in

Atom	*Type*	δ
SIA-C2	*CG031*	*0.2e*
SIA-O6	*OG301*	*-0.31e*
GAL-C6	*CG321*	*0.3e*

Table 5: Parameters assigned to the atoms of the sialic acid-galactose bond. The atom type refers to types from the Carbon generalized force field (CGenFF) [166] and δ is the assigned partial charge.

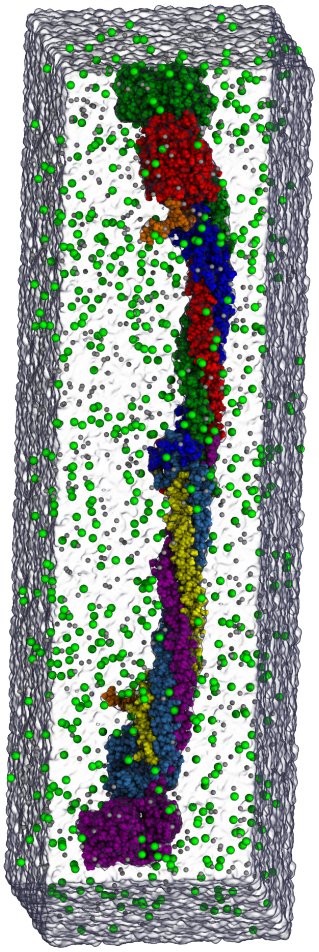

Figure 15: Initial simulation box of the fibrinogen dimer.

a crystal structure. A partial structure of the bovine αC-region is available in the PDB entries 2BAF and 2JOR and shows the β-hairpin structure that is stabilized at its base by a disulfide bond. The missing 170 residues between the end of the Aα-chain in 3GHG and the beginning of the αC-region in 2BAF make it unfeasible to predict the structure for the intervening region computationally. Additionally, the size and flexibility of the αC-region would prevent an accurate sampling of its behavior with current computational techniques. For this reason the αC-region is neglected in the simulations despite its likely interaction with the fibrinogen core and its relevance for intermolecular interactions [21, 22]. The αC-region will likely influence adsorption and especially aggregation properties of fibrinogen so the simulation results starting from the described structure should be treated with care.

Another point that should be mentioned are the systematically missing residues in the central domain of fibrinogen. These correspond to further flexible parts that include the fibrinopeptides. Therefore, the simulations can not distinguish between the behavior of fibrinogen and fibrin monomers. Two additional carbohydrate cluster located in the coiled-coil region (attached to residue γ52) are not resolved in the crystal structure and have not been included in the simulation.

Overall 477 residues are missing from the crystal structure, while 956 are resolved. A slight structural difference between the two protomers in the selected structure is present. The unfolded region in the γ-chain at the center of the coiled-coil region is larger in one protomer than the other. During the simulation this region does not refold into a helical state.

Topology and Carbohydrate Parametrization

The structure of the peptide chains was build using the psfgen utility of VMD and the CHARMM27 topology [161]. The protonation state of the amino acids was manually set to fit a pH of 7. At this pH level, only histidine can have two protonation states. One, where the hydrogen is attached to the ε-nitrogen and one, were it is attached to the δ-nitrogen. The appropriate state was determined by visual inspection in such a way as to give the best balance between solvent contact and the possibility for the formation of intra-molecular hydrogen bonds. The position of hydrogen atoms for all residues was then determined based on the CHARMM topology as they are not resolved in the X-ray structure.

Disulfide bonds were added corresponding to the definition in the PDB file. This leaves no free cystein residues and connects the different peptide chains. In the following, dimer- as well as protomer-systems of fibrinogen were build to study effects of dimerization. No artificial constraints on the structure were employed for the protomer simulations as the N-terminal disulfide ring provides enough stability for the coiled-coil region.

The topology for the sugar groups was build using the recent extension of the CHARMM force field to carbohydrates [167]. At the time the structures were build, the bond connecting the terminal sialic acid residue to the preceding galactose residue was not yet parametrized. Since then, an improved version of the force field was published [168], but has not been used in the simulations.

To determine the parameters for the bond, all involved atoms were assigned an atom type using paramchem [259] (see Table 5). This is a web interface that assigns atom types from the CHARMM carbon generalized force field (CGenFF) [166] to small organic molecules based on their structure. The full galactose – sialic acid compound was used for this. The penalty scores provided by paramchem indicated a good fit to existing atom types. The partial charges varied slightly between the two structures in the biantennary carbohydrate cluster. Within the limits provided by these structures, the assigned partial charges where adjusted to ensure charge conservation upon bond formation. Both protomer and dimer structures were build with and without glycosilation.

Simulation Boxes and Protocol

A water box was created to solvate the protein for simulation (see Fig. 15 and Fig. 16). This is necessary to reproduce realistic behavior in the simulation. The rectangular simulation boxes were build with explicit TIP3P water [162] and a physiological ion concentration of 15mMol NaCl using VMD [260]. The size of the resulting boxes are given in Table 6. Periodic boundary conditions were used in all directions to mimic an infinite water box. Thus, the behavior of fibrinogen in dilute solutions is studied in these simulations.

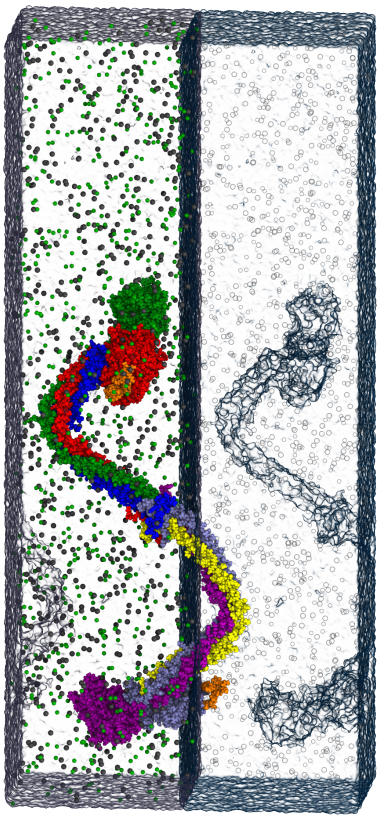

Figure 16: Simulation box of the fibrinogen dimer after 77ns. The deformations lead to part of the molecule sticking into the periodic image box.

The water box is build in such a way as to take advantage of the elongated structure of fibrinogen. For this reason measures have to be taken to keep fibrinogen aligned to the long axis of the box without influencing its dynamics. This is achieved using collective variable constraints. To estimate the effect of the collective variable constraints, some unconstrained simulations were performed. A discussion of the constraints used and their effect on fibrinogen dynamics is given in section 4.3.

The interactions were provided by the CHARMM27 force field with CMAP corrections [161, 163] and its extensions to sugars and other organic molecules discussed in section 3.2 [166, 167]. All simulations were performed in an NPT ensemble at T=310K and P=1atm using NAMD [261] with a Langevin thermostat and a Langevin piston barostat. Other settings as well as the thermostat and barostat are described in section 4.1.3.

The simulation boxes were first subjected to 5000 steps of energy minimization and then equilibrated. For the minimization the heavy protein atoms were fixed to their positions in the crystal structure while the water was minimized. After this mini-

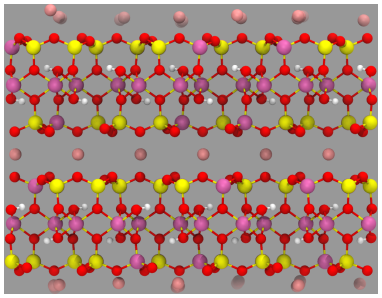

Figure 17: Sheet structure of mica. Silicone atoms are shown in yellow, aluminum in purple, potassium ions in pink and oxygen and hydrogen in red and white respectively.

System	Initial box size [nm]	N	Simulation time [ns]
Dimer, glycosilated	13.27 × 48.59 × 12.70	788173	77 + 88
Dimer, unglycosilated		786811	25 + 20
Protomer, glycosilated	12.28 × 27.89 × 11.57	381397	199 + 188
Protomer, unglycosilated		380169	135 + 109 + 100 + 82 + 51 + 43 + 30 + 20 + 20 + 14

Table 6: Initial simulation box sizes, particle numbers and length of the trajectories for the different systems of fibrinogen.

mization, MD runs were performed over a timespan of 1.6ns (protomer) and 2ns (dimer). During these runs the constrains on the heavy atoms were gradually lowered from 3kcal/mol to 0kcal/mol and the temperature increased from 110K to 310K. Following the removal of the constraints the systems was allowed 1ons of unconstrained simulation time to equilibrate.

Various simulation runs with different initial velocities were started. The length and number of the production runs are given in Table 6.

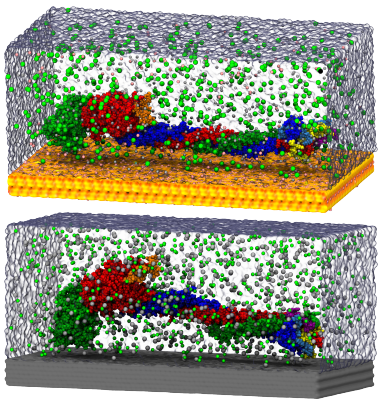

Figure 18: Simulation boxes of the fibrinogen protomer at a mica (top) and graphite (bottom) surface.

4.1.2 *Simulation Set Up and Preparation for Fibrinogen at Interfaces*

Mica and graphite were chosen as model surfaces to investigate the adsorption behavior of fibrinogen. Graphite is a model hydrophobic surface without any long range interactions while mica is a hydrophilic surface with long range electrostatic interactions. Additionally, both are used frequently in adsorption experiments [17, 127, 129, 130, 132, 142, 143] because their sheet structure allows the creation of atomically flat surfaces. Both surfaces have previously been used for protein adsorption simulations [225, 227, 233–236, 238]. The atomistic resolution was chosen because it is known that the interaction of the protein and surface hydration layers can influence adsorption [228, 229, 262].

Structure of the Surfaces

As graphite is a well known material with a simple surface chemistry, it will not be discussed here. For the adsorption the graphite surface was prepared as a 6 layer graphene sheet using the VMD Carbon Nanostructure Builder. The carbons were assigned

the C_α-atom type from the CHARMM27 force field.

Mica on the other hand is more complex and needs more description. Mica is a sheet silicate with the formula

$$K_{1.0}[Si_3Al_1O_8][Al_2O_2(OH)_2]$$

and belongs to the clays. It consists of layers of silicon aluminum oxide separated by layers of potassium ions. The structure is shown in Fig. 17. A structure for a mica unit cell and a super cell with realistic defect distribution, as well as force field parameters [170] were kindly provided by H. Heinz (University of Akron). The defects are point defects where an aluminum atom is substituted for a silicon atom. The structures were integrated into the inorganic builder of VMD. In collaboration with the group of H. Heinz, the force field parameters were transfered from LAMMPS to CHARMM format and tested. In contrast to other force fields for silicates [169], the mica force field reproduces surface and bulk properties.

A two layered mica surface was used for the simulation. The provided structure places all potassium ions on one side of the surface. This leads to a large dipole moment across the simulation box that is much stronger than what is present under experimental conditions. To achieve a realistic ion distribution, a slab of mica was simulated in vacuum under periodic boundary conditions. The simulation was started at 1350K, that is 50K above the melting point of mica, with the electrostatic interaction turned off. This lead to a uniform distribution of the potassium ions in the vacuum. Reactivating the electrostatic interaction and repeatedly running short stretches of MD simulation interspersed with energy minimizations while reducing the temperature led to a mica slab with evenly distributed potassium ions across its two sides. The resulting arrangement of potassium ions is shown in Fig. 19. Due to the lack of covalent bonding of the surface potassium ions, they will be able to diffuse into solution during the simulation.

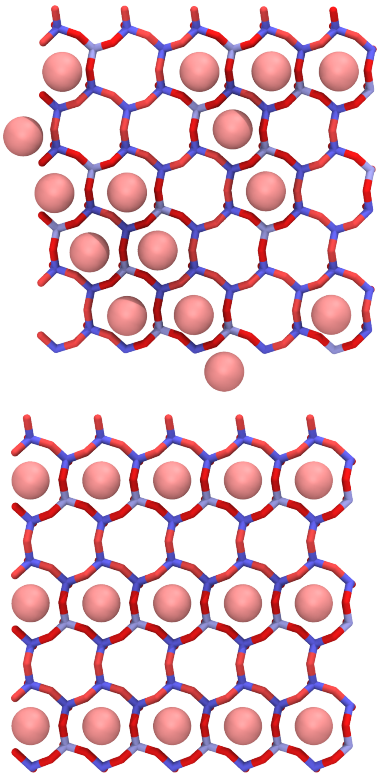

Figure 19: Charge structure (red negative to blue positive) and distribution of potassium ions on a mica surface with defects (top) used for adsorption simulations and a defect free surface (bottom).

Simulation Box and Protocol

The simulation set-up was performed accroding to established procedures for protein adsorption simulations discussed in section 3.4. The simulation box for the surface system is periodic in all dimensions. However, the surface will be continuous in the xy-plane. This is enabled by defining covalent bonds that correctly wrap around the periodic boundaries. On top of that it is necessary to achieve a constant pressure simulation by allowing box fluctuations only in the z-direction. In this way, the area of the xy-plane is constant and an infinitely continuous surface is approximated.

Each surface was surrounded by an appropriate water box with physiological salt concentration (see Fig. 18) and the energy was minimized. The system was then equilibrated by slowly

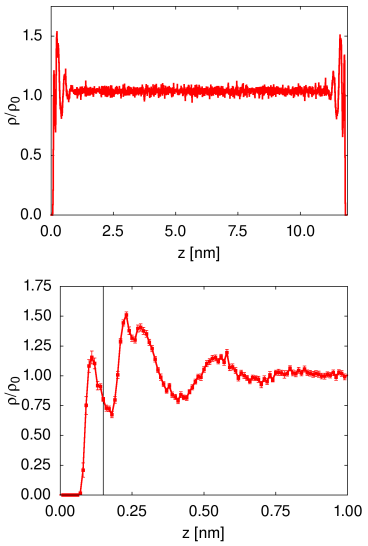

Figure 20: Density of water oxygen atoms in the mica simulation box. The black line denotes the position of the potassium counter ion layer.

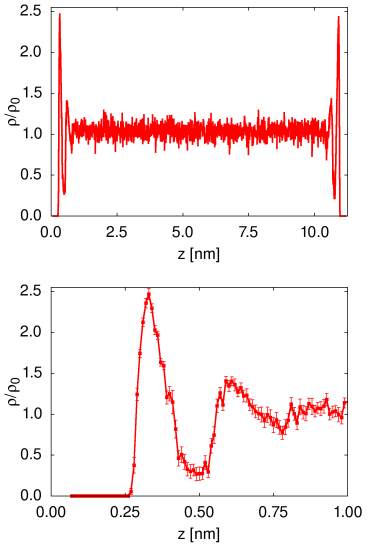

Figure 21: Density of water oxygen atoms in the graphite simulation box.

releasing constraints on the surface atoms at a temperature of 310K. In a subsequent step, the system was simulated for 0.6ns before the first protomer and residues α27-65,β58-95 and γ14-40 of the second protomer were added. The structure of the protomer was taken from the end of the 10ns equilibration run of the dimer solution system. The residues of the second protomer were included to complete the E-region.

It is not feasible to simulate systems in which fibrinogen can leave the surface, reorient itself in solution and adsorb in a more favorable orientation. This is due to the large system sizes and timescales involved in such simulations. To avoid such simulations, an orientational screening was performed. The protomer was placed with its long axis parallel to the surface and rotated around this axis to create three different orientations. This procedure limits the sampled space to so called side-on adsorption which is known to be the case, at least for dilute solutions [17, 255]. The orientations are labeled after the rotation angle 0, 120, and 240 respectively in the following. The orientation 0 is positioned in such a way that the carbohydrate cluster points away from the surface. For orientation 240 the carbohydrate cluster points toward the surface and is located between surface and coiled-coil region. The orientation 120 is intermediate and has a side on orientation in the D-region with the P1 side pointing away from the surface. Due to the orientation algorithm, the different orientations on mica and graphite are similar to each other but vary by a few degrees. In each orientation, fibrinogen had a minimal distance of 0.8nm to the surface. This distance was measured with respect to the potassium counter ion layer in the case of the mica surface. Water molecules intersecting the protomer were removed and the ion concentration adjusted to compensate for the charges on fibrinogen. The dimensions of the surface systems are given in Table 7.

The heavy protein and surface atoms were again fixed while the solvent was minimized. The constraints were gradually released over a 0.75ns equilibration period. After that, several NPT MD simulations were performed with NAMD. For the adsorption simulation no collective variable constraints were used to keep the protein aligned to the long axis of the box. Instead, a steeply increasing harmonic potential was used in z-direction starting 10nm above the surface to prevent adsorption at the periodic image surface. This force was implemented using the tclForces module of NAMD. The lower sheet in mica and the lowest two sheets in graphite were restrained to their initial positions using a 3kcal/mol harmonic restraint. This is done to mimic the effect of bulk material.

The application of a barorstat exclusively in z-direction together with constrained surface atoms might lead to artificial pressure fluctuations. To verify that this is not the case, the water density profile was calculated for a crossection of the simulation box not containing parts of the protein. The resulting profile was

System	Initial box size [nm]	N	Simulation time [ns]
Graphite, 0°		440893	51 + 48 + 45 + 20 + 20 + 18
Graphite, 120°	12.18× 27.36 × 13.03	440875	52 + 50 + 49 + 49 + 26 + 19
Graphite, 240°		440698	51 + 49 + 48 + 44 + 26 + 23
Mica, 0°		681900	120 + 120 + 95 + 94 + 71 + 70 + 48 + 48 + 45 + 43
Mica, 120°	18.47 × 29.91 × 14.30	681937	120 + 96 + 91 + 90 + 85 + 70 + 45 + 45 + 44 + 43
Mica, 240°		681909	120 + 100 + 70 + 70 + 70 + 46 + 45 + 41 + 38 + 30

Table 7: Initial simulation box sizes, particle numbers and length of the trajectories for the different systems of fibrinogen.

averaged over a 0.5ns stretch of simulation time. The results for mica and graphite are shown in Fig. 20 and Fig. 21. The observed density profiles are in accordance with known water density profiles at silicates [263] and graphite [264].

4.1.3 *Molecular Dynamics simulation with NAMD*

All atomistic simulations were performed using NAMD [261]. NAMD is a program to efficiently perform MD simulations of biological systems on parallel computers.

Parallelization and Time Integration

To achieve its performance the simulation box is divided into different grid cells that are then distributed to different CPUs (domain decomposition). If a particle crosses the boundaries of these boxes, or if long range interactions are evaluated, communication between the CPUs is necessary. If this happens to often, the time invested in communication will take up more of the total computation time than is gained by distributing the

calculation to more CPUs. Due to this slow down, the simulation speed will not scale linearly with the number of processors and the speed up will dimish as the system each CPU simulates gets smaller and smaller. An example of such scaling behavior is shown in Fig. 22.

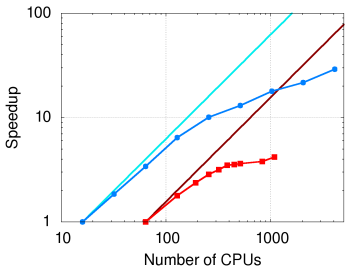

Figure 22: Speedup of NAMD on the MOGON (Mainz, red) and the HERMIT (Stuttgart, blue) clusters. The speedup is calculated with respect to a single node. The straight lines represent ideal scaling behavior.

The evaluation of forces was just mentioned as one of the factors influencing the speed with which a simulation can be performed. As forces become longer ranged, it becomes more costly to evaluate them. On the other hand the long range part is less sensitive to small scale changes. This fact is used by NAMD by employing a multiple time step algorithm. The bonded forces are evaluated most frequently (in every time step). The van der Waals forces and the short ranged part of the electrostatics are evaluated only every n-th step. The long ranged electrostatic interactions are only updated every m-th step where $m > n$. In this work, the time step was set to 2fs and the van der Waals interaction were also evaluated every time step. The long ranged electrostatic interactions were updated every second step. These parameters where chosen to avoid non-linear resonances that can build up in this kind of integrator [265] and lead to instability of the simulation. To further increase the speed of the simulation, the bond lengths of hydrogen bonds to heavy atoms was fixed. This eliminates high frequency vibrations that would reduce the possible time step.

Non-bonded interactions and the Particle Mesh Ewald Method

The van der Waals interactions are short ranged non-bonded interactions, i. e. they can be treated by introducing a cutoff for the potential. This means only atoms in the neighborhood of an atom contribute to the forces on it. The atoms in the neighborhood can be tracked during the simulation so that the van der Waals forces can be evaluated very fast. The cutoff was set to 1.2nm and the potential was smoothly reduced to zero after 1nm to avoid diverging forces at the cutoff. Such cutoff based methods only work for short ranged interactions that got to zero more rapidly than r^{-3} as $r \to \infty$ [148, ch. 12]. The Lennard-Jones potential

$$V_{VdW}(r) = \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{4.1}$$

describing the van der Waals interaction is a prime example of this.

To deal with long range interactions, such as the r^{-1} Coulomb potential describing electrostatics, one needs to find a method that can evaluate the forces cheaply without a cutoff. One way to do this is the so called Ewald summation that takes advantage of the periodic boundary conditions in the simulation. In this approach, the interaction is split into a well behaved long range part and a short range potential that quickly drops to zero but contains singularities. Such singularities are caused, e.g. by

point particles. The long range potential is then free of singularities and can be evaluated in Fourier space. The summation in Fourier space can be made more efficient by a wavenumber cutoff, amounting to a short range cutoff in real space below which the contributions of the long range potential are not counted. The short range potential will fill these contributions in and can be treated in real space. The use of the Fourier transform implies an infinitely periodic systems. Due to the periodic boundary conditions, this is usually the case in simulations. The potential energy in the Ewald summation is based on the solution of Poisson's equation for the charge distribution determined by the position of the point charges in the simulation. This solution can be made more efficient by placing the charges on a mesh. In NAMD, this is done using the Smooth Particle Mesh Ewald (SPME) method, where B-splines are used to perform an interpolation that distributes the charges on the nodes of the mesh [261].

The Langevin Thermostat and Langevin Piston Barostat

As the simulations were performed in an NPT ensemble, the temperature and pressure need to be kept constant. This means energy and volume fluctuations according to the Boltzmann distribution have to be introduced into the system.

The temperature fluctuations were introduced by performing Langevin dynamics simulations. In this approach, a random white noise force $\mathbf{R}(t)$ is added to the physical forces described earlier. Just adding the random force would continuously pump energy into the system and lead to instabilities. To remove energy from the system, a dissipative term is added. The prefactors are chosen so that the fluctuation-dissipation theorem is fulfilled. The resulting equations of motion are:

$$m\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) - \gamma\dot{\mathbf{r}} + \sqrt{\frac{2\gamma k_B T}{m}}\mathbf{R}(t)\,. \tag{4.2}$$

The damping coefficient was chosen as $\gamma = 1\mathrm{ps}^{-1}$.

The volume fluctuations necessary for constant pressure are obtained by adjusting the size of the unit cell and then rescaling the atomic coordinates accordingly. To this end, the system is viewed as a piston that can expand and contract depending on the values of the current internal pressure and the reference (piston) pressure. The extension of the piston can be integrated into the equations of motions using an extended Lagrangian approach [266], the so called Nosé-Hoover barostat. As the piston has a mass in this approach, its inertia leads to unphysical "ringing" effects with a period τ_P. To minimize these effects, the piston degrees of freedom are not evolved freely but undergo Langevin dynamics [267]. The resulting method is called Nosé-Hoover Langevin Piston barostat. The parameters chosen for the

fibrinogen simulations are $\tau_P = 200fs$ and the friction for the piston was chosen such that damping timescale was $100fs$.

4.2 ANALYSIS METHODS

The simulation trajectories provide a wealth of information. To elucidate the complex behavior of fibrinogen based on these trajectories a variety of analysis techniques is necessary. Simple tasks like visualization, contact, distance and angle calculations as well as root mean square deviation (RMSD)

$$\mathrm{RMSD}(t) = \left(\frac{1}{N} \sum_{i=1}^{N} |\mathbf{r}_i(t) - \mathbf{r}_{i,ref}|^2 \right)^{\frac{1}{2}}, \tag{4.3}$$

and root mean square fluctuation (RMSF)

$$\mathrm{RMSF}_i = \left(\frac{1}{T} \sum_{n_t=1}^{T} |\mathbf{r}_i(n_t) - \mathbf{r}_{i,ref}|^2 \right)^{\frac{1}{2}} \tag{4.4}$$

calculations were performed with VMD [260]. For the RMSD and RMSF calculations only the C_α atoms were selected.

Other techniques for more specialized analysis are presented in the following along with the programs used.

4.2.1 *Principle Component Analysis*

One of the most important behaviors to analyze is the motion of fibrinogen and its different parts. A straightforward way to estimate the motion of proteins is the so called normal mode analysis. In this method, the energy landscape of a certain protein structure is calculated. Based on this, a harmonic approximation around the minimum can be used to determine the modes of motion. Such an analysis can be performed relatively easy given a suitable energy function as only a single protein structure is required. The drawback is of course that only harmonic motions are considered. An improved way to characterize the motions of proteins is to perform a Principle Component Analysis (PCA) [268]. In this method, the covariance matrix

$$C_{ij} = \left\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \right\rangle \tag{4.5}$$

is calculated by performing the average over an MD trajectory and is thus more costly than a normal mode analysis. Diagonalizing this matrix gives the principle components of motion of the protein as the eigenvectors of the covariance matrix. These modes account for the variance between residue positions according to their eigenvalue. The eigenvalue associated with a given component correspond to the variance in this direction of motion and the different directions of motion are per definition linearly uncorrelated. The eigenvalue spectrum of proteins usu-

ally contains only few eigenvalues that contribute significantly to the overall motion. This is illustrated by the fact that typically less than 1% of the modes account for over 90% of the fluctuations [269]. The subspace responsible for the dominant, long time scale motions is called essential dynamics (ED) subspace.

A PCA can also be used to judge if the observed motions are likely biologically relevant, i.e. represent movements between different free energy minima, or just random diffusion within one minimum. The later is the case if the first principle components show a cosine like behavior that is associated with the free diffusion in an energy minimum [270].

Application to fibrinogen simulations

To study the relevant dynamics of (parts) of fibrinogen a PCA is employed focusing on the motion of the C_α-atoms. The programs wordom [271] and ProDy [272] are used to perform the PCA.

In a first step, the C_α-atoms are extracted from the simulation trajectory and translational and rotational modes are eliminated by aligning the structures in each frame to the initial structure. The correlation matrix is then calculated from the aligned trajectory of the C_α-atoms and the eigenvectors and eigenvalues are determined. The eigenvectors are sorted according to their associated eigenvalues. The dominant ten eigenvectors are used for analysis and their projection along the simulation trajectory is calculated to determine the correlation with other quantities. For the analysis of the large scale behavior the four longest protomer trajectories were mergeded (see Table 7, 631ns of cummulative simulation time).

The similarity between the motions in different parts of the trajectory or between different trajectories was determined by an overlap between the respective essential dynamics subspaces [270]. This overlap is defined as

$$O(\{\underline{x}_i\},\{\underline{y}_i\}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\underline{x}_i \cdot \underline{y}_j\right)^2 , \tag{4.6}$$

where $\{\underline{x}_i\}$ and $\{\underline{y}_j\}$ are the two subspaces spanned by principle components. The following equations hold:

$$\{\underline{x}_i\} \cap \{\underline{y}_i\} = \emptyset \quad \Leftrightarrow \quad O(\{\underline{x}_i\},\{\underline{y}_i\}) = 0 \tag{4.7a}$$

$$\{\underline{x}_i\} = \{\underline{y}_i\} \quad \Leftrightarrow \quad O(\{\underline{x}_i\},\{\underline{y}_i\}) = 1 \tag{4.7b}$$

For the overlap calculations the PCA Analysis was performed on merged trajectories of comparable size, to avoid a bias to small scale movements. For a comparison of different collective variables (see section 4.3) only stretches of 40ns after the start of the simulation were compared. To compare protomer and dimer motions, protomer trajectories were extracted from the simulations of the glycosilated dimer. The four resulting trajectories

were merged to perform a PCA that yields modes that can be compared to the modes obtained from the protomer simulations.

4.2.2 *Linear Mutual Information and Network Analysis*

Linear Mutual Information as a measure of correlation

To improve the analysis of correlated motions in proteins it was suggested to use the mutual information instead of the Pearson correlation in eq. (4.5) as a basis for analysis [273]. The mutual information for variables $\mathbf{x}_{1,\dots,N}$ is defined as

$$I(\mathbf{x}_1,\dots,\mathbf{x}_N) = \int d^3x\, p(\mathbf{x}) \ln\left[\frac{p(\mathbf{x})}{\prod_{i=1}^{N} p_i(\mathbf{x}_i)}\right] . \tag{4.8}$$

This quantity can be related to a generalized correlation coefficient $r_{MI}^{(\mathbf{x}_i,\mathbf{x}_j)}$ that describes how well the best non-linear model can relate $\mathbf{x}_i$ to $\mathbf{x}_j$. It is given by

$$r_{MI}^{(\mathbf{x}_i,\mathbf{x}_j)} = \left(1 - e^{-2I[\mathbf{x}_i,\mathbf{x}_j]/d}\right)^{-\frac{1}{2}} , \tag{4.9}$$

where d is the dimensionality of the problem. The generalized coefficient r_{MI} is zero for uncorrelated motions and one for fully correlated motions. Despite its usefulness in identifying correlated regions in proteins [273, 274], the estimation of the mutual information is challenging and demands high accuracy to get an accurate estimate for r_{MI}. The determination of modes of motion based on the mutual information, called full correlation analysis, is also computationally very costly [275]. Therefore, the linearized version of the mutual information (LMI) will be used. It can be calculated from the correlation matrices of the variables according to

$$LMI(\mathbf{x}_i,\mathbf{x}_j) = \frac{1}{2}\left(\ln\left[\det C_i\right] + \ln\left[\det C_j\right] - \ln\left[\det C_{ij}\right]\right) \tag{4.10}$$

with

$$C_i = \left\langle \mathbf{x}_i^T \mathbf{x}_i \right\rangle \tag{4.11}$$

and C_{ij} given by eq. (4.5). Here the variable $\mathbf{x}_i$ correspond to the position of the C_α-atom of residue i in 3-dimensional space and the average is taken over the simulation trajectory.

Computational detection of allosteric pathways

Allostery, or allosteric regulation, is a regulation mechanism in proteins. In this mechanism the affinity of one binding site for its ligand is influenced by the binding of a ligand to a different site. As allostery can be a very subtle effect and the affected sites can be seperated by substantial distances. Therefore, it is difficult to detect in computer simulations. It is often not possible to

simulate the binding of multiple ligands to observe changes in the binding site affinity so that experimental knowledge about the existence of an allosteric effect is usually required.

To identify the mechanism responsible for allostery, most computational approaches rely on constructing a network that represents connections among residues. This network can be based on residue contacts, correlations between residues, residue interactions, or dynamics [274, 276–279]. The networks based on dynamics are so called elastic network models of the protein that were discussed in more detail in section 3.3.2.

Once the network representing the protein is constructed, the analysis can be taken into two directions [279]: The first is to identify pathways between the two binding sites by traveling along connected nodes in the network. Special attention is paid to the shortest path that can be found in such an analysis as the signal would be attenuated the least. The distance between network nodes can be adjusted according to their interaction strength or correlation to account for this signal attenuation, alternatively the topology of the network can be adjusted by removing connections if the correlation is below a certain threshold. The second method would be to introduce perturbations in the network and then monitor the propagation of the signal from the perturbation. Such a perturbation can be introduced either by moving one residue with a periodic driving force or by introducing "point mutations" in the network by modifying connection strength. Such network analyses can also be used to compare homologous proteins if it is known that, e.g. one shows an allosteric effect while the other does not [278]. The difference in the network can be used to identify residues crucial for the allosteric behavior. It has been suggested that such crucial residues can also be identified in another way [277]: By constructing the network so called hubs are identified. These are residues that are highly connected so that a signal can reach these residues from many points.

In this work, a network pathway analysis based on LMI correlation of residues will be used. The LMI correlation will be used as a recent study has shown that such generalized correlation measures are better suited to detect allosteric pathways than standard Pearson correlation coefficients [274].

To identify communication pathways between different residues of the protein, a network approach based on the LMI correlation matrix is taken. The LMI correlation matrix is calculated for glycosilated and unglycosilated simulations separately, to identify effects resulting from glycosilation. Based on this correlation matrix a network can be constructed in which each node represents a residue. The nodes are connected if their LMI correlation exceeds r_{crit}. All connections have the same length. Alternatively the network can be constructed from other data, e.g. the hydrogen bond network. The most significant communication between two residues is assumed to happen via the shortest

high correlation path connecting them in the constructed network. Therefore, only the shortest connection between two given nodes was determined by a recursive search. Because the search space can become very large if the network contains many nodes and/or if the correlation matrix is not sparse, a simple search is too slow to yield results.

One way to improve the search time is to restrict the search to paths of a maximum length N. This length was determined by the criterion

$$(r_{crit})^N = 0.025\,, \tag{4.12}$$

i.e. the maximum path length was such that the connected residues had a correlation of at least 2.5% along the path. If a path connecting start and target node with less than N steps was found, N was reduced to this number. Therefore, only paths of minimal length can be reliably identified.

A second way to improve the speed of the algorithm is pruning of the searched network. In a first step, all closed loops where removed from the network. In a second step, for each recursion level all nodes were saved for which no path to the target node can be found starting from the current level. Further recursions check possible paths against this list and ignore all paths that start with a node that is known to have no connection to the target node.

With those improvements, an identification of pathways between nodes in networks with ≈400 nodes is possible in less than one minute on standard Desktop Computer. This is sufficient for the purposes of this work.

4.2.3 *Cluster Analysis*

To identify important conformational states, the conformations sampled by the MD simulation were clustered. To this end, the "Leader-like" clustering algorithm in wordom [271] was used. In this algorithm, a new cluster is defined by a conformation in the simulation that is sufficiently far apart from previous clusters according to some measure. The clustering is performed by a single run through the trajectory and no iterative refinement of the clusters is done.

As a measure of distance, the root mean square deviation of internal distances (distance root mean square, DRMS) is used. To find an appropriate cutoff, the DRMS is first calculated for the entire trajectory and then a histogram is made. The cutoff is set in the center of the first minimum after the peak corresponding to local fluctuations.

4.2.4 *Fourier Analysis*

To determine the time scale of periodic motions, a Fourier Analysis is employed to determine relevant frequencies. The motion is characterized by the fluctuation in a characteristic distance. This distance is monitored during the simulation. The Fourier spectrum of the distance fluctuations is calculated using the discrete cosine transform of Mathematica6. This spectrum was calculated for each trajectory individually and later averaged.

As proteins have scale free dynamics [198], the spectrum should show a power law behavior and functionally relevant motions manifest as deviations from this behavior. The deviations are approximated by a Gaussian on top of the background power law behavior.

4.2.5 *Secondary structure and H-bond analysis*

As described in section 2.1.1, the secondary structure is an important feature of protein structure. To standardize the assignment of the various secondary structure elements, Kabsch and Sander [280] developed a program that uses the backbone hydrogen bonding pattern in a 3D protein structure. In this program, a hydrogen bond is defined electrostatically by assigning a partial charge of $0.20e$ to the donating group and $-0.42e$ to the acceptor group. A hydrogen bond exists if the resulting interaction energy is less than $-0.5\mathrm{kcal/mol}$. Only the two strongest hydrogen bonds are counted for each residue. These hydrogen bonds can form simple patterns such as an n-turn where a hydrogen bond is formed between residue i and $i+n$ or a bridge where non consecutive parts of the protein are connected. Consecutive turns and bridges build the known secondary structure elements like helices and β-sheets (see Table 8). Additionally, bends are identified as high curvature regions, where the curvature is measured along the protein backbone. Regions that can not be assigned to any of the groups are designated as coils. As the assignment is based on combining smaller units into larger ones, it is relatively resistant against small defects in the hydrogen bonding pattern. The assignment of the secondary structure was performed with a recent implementation of the DSSP algorithm [281].

Element	*Definition*
3_{10}-*helix*	*consecutive 3-turns*
α-*helix*	*consecutive 4-turns*
π-*helix*	*consecutive 5-turns*
β-*ladder*	*consecutive bridges*
β-*sheet*	*consecutive bridges connected by shared residues*
bend	*high backbone curvature*
coil	*no pattern*

Table 8: Secondary structure elements defined in DSSP

The basic idea of DSSP highlights the importance of hydrogen bonds for protein structure. Additionally, it is known that hydrogen bonds can be crucial for the functioning of proteins, e.g. in allosteric regulation [278] or ligand binding [24]. Therefore, all possible hydrogen bonds in fibrinogen were monitored along the trajectories. This was done using wordom [271] identifying a hydrogen bond not using the electrostatic interaction energy but rather by geometric features. A hydrogen bond is identified

if the donor and acceptor groups are closer than 3.1Å and the donor – hydrogen – acceptor angle is less than 130°.

4.2.6 *Dyndom*

If rigid domains in a protein are to perform large relative motions with respect to each other, they need to be connected by flexible linkers. Such flexibility can, e.g be delocalized over a long flexible region. Another possibility is the presence of a very localized hinge interrupting rigid regions. To identify such localized hinges, the program dyndom [282] was used. This program works by analyzing the two extremal structures of the relative domain motion. For each atom the rotation is calculated that would be needed to bring an atom from its positions in one structure to its position in the other. A clustering of these rotations is performed. The resulting clusters represent atoms that move together and likely form a rigid domain. From the relative motion of the rigid domains the connecting hinges can be determined. Residues are considered to be hinge residues if they are at the interface of two domains for which the inter domain motion exceeds the intra-domain motion. The relative motion is then described by a screw motion. If the axis of this screw motion is located close to the inter-domain residues, these residues are identified as a mechanical hinge.

For the analysis of the large scale motion of fibrinogen with dyndom structures were used that provide the maximum and minimum PCA projection with respect to the dominant PCA component.

4.2.7 *Electrostatics*

The electrostatic properties of fibrinogen are characterized by the electrostatic potential they create. The potential that is created by the charges in an electrolyte solution can be calculated by solving the Poisson-Boltzmann equation:

$$\boldsymbol{\nabla}\varepsilon(\mathbf{r})\cdot\boldsymbol{\nabla}\Phi(\mathbf{r}) = -4\pi\left(\sum_N Q_N\delta(\mathbf{r}-\mathbf{r}_N) + \sum_n \frac{q_n e^{-\frac{q_n\Phi(\mathbf{r})}{k_B T}}}{\int dV\, e^{-\frac{q_n\Phi(\mathbf{r})}{k_B T}}}\right), \tag{4.13}$$

where ε is the permittivity, Φ is the electrostatic potential and Q_n and q_n are the charges of protein atoms and salt ions in the system respectively. The Poisson-Boltzmann equation is a nonlinear partial differential equation for the electrostatic potential of an electrolyte solution. In this description, the electrostatic potential is determined by the charge density via Gauss's law. On the other hand the charge density of the system is given by a statistical mechanics description of the system of mobile ions in solution. This description of course relies on the knowledge

of the electrostatic potential to determine the correct Boltzmann factors. The equation can be linearized for the case $q_n\Phi \ll k_B T$:

$$\nabla\varepsilon(\mathbf{r})\cdot\nabla\Phi(\mathbf{r}) = -4\pi\left(\sum_N Q_N\delta(\mathbf{r}-\mathbf{r}_N) + \sum_n q_n\left[1-\frac{q_n\Phi(\mathbf{r})}{k_B T}\right]\right). \tag{4.14}$$

The linearized Poisson-Boltzmann equation was solved using the APBS software [283] with a multi-grid scheme [284] and zero boundary conditions. The periodic bounadry conditions used in the simulation were not used in these calculations. The grid size was chosen so that it extended at least 6nm beyond each protein atom. Zero boundary conditions were used for the boundary of this grid because the salt solution should provide enough screening for the protein charges at such distances. The results are the tabulated values of $\Phi(\mathbf{r})$ at the grid points. To take thermal motions of the protein charges into account, the potential was calculated for the simulation snapshots spanning 0.25ns and the results on the grid points were averaged.

4.2.8 *Sequence Comparisons and Analysis*

To estimate the significance of features identified in the simulation, an evolutionary comparison is performed. If a feature is relevant to the correct physiological functioning of a protein, the sequence of amino acids leading to this feature should be highly conserved.

Fibrinogen sequences of different species where identified by searching the UniProt data bases Swiss-Prot and TrEMBL [285, 286] for protein names containing "fibrinogen" or proteins coded for by the genes FGA, FGB or FGG. Species were only considered if a γ-chain fragment was present that contained enough residues after the N-terminal disulfide ring to extend to the identified hinge region. In this way, γ-chain fragments from 83 species were identified. Of these the full sequence for all three fibrinogen chains was available for 33 species. The exact list of species as well as their evolutionary relationship can be found in Appendix A. The lamprey sequence contained in the set belongs to the jawless vertebrates and forms the basal clade of the vertebrate tree. Having branched of earliest from the main vertebrate lineages, lampreys also have the simplest clotting process of all vertebrates [31]. Therefore, deviations from the other considered species are expected.

The sequences were aligned using the Chimera MatchMaker utility [287, 288] with a Needleman-Wunsch alignment algorithm and BLOSUM62 matrix to score the alignment. The sequences were aligned to the human sequences. The alignment predictably aligns the disulfide rings in all species, so that the amino acids following them will be in the same position along the coiled-

coil region for all species. Based on this alignment a phylogenetic tree for the γ-chains was created using the UniProt website. Furthermore, the alignment allows to judge the evolutionary conservation among vertebrates. Examples are discussed in sections 5.2.2 and 5.4.2.

From the aligned sequences the regions corresponding to the coiled-coil regions where extracted. These regions where identified by the position of the disulfide rings [25]. The result are three 120 residue long sequences for each fibrinogen that are in register (see Table 9), i.e. if counted from the disulfide ring, the same residue number corresponds to the same position along the coiled-coils for each sequence.

Chain	*Sequence*	*Residue numbers*
α	`CPSGC...CRGSC`	*45–165*
β	`CPTGC...CRTPC`	*76–196*
γ	`CPTTC...CQEPC`	*19–139*

Table 9: Definition of coiled-coil sequences in human fibrinogen

Simulations of different fibrinogens from a representative group of species is out of the question (partly due to the computational cost and partly due to the unknown structure as a basis for the simulation). Despite the lack of experimentally determined structures and simulations, it is possible to gain information about key structural features of the protein. This is possible by using heuristic techniques that predict structural features based on the primary sequence.

One key feature is the helix propensity or the predicted helix content of a certain amino acid sequence. This can be estimated using the agadir program [289–291]. This program is based on treating a polypeptide as a set of two state systems, one state being a helical conformation and the other being a random coil. Different parameters are needed to describe the helix coil transition. These parameters include, e.g. the intrinsic affinity of different amino acids for helical conformations or the hydrogen bonding with their neighbors. They can be determined with the help of experiments. In this approach, a sequence has to have at least four amino acids which are assigned a helical conformation before an α-helix is detected [289]. Nowadays refined parameters are available and the model was extended to include electrostatic and ion effects [290]. The agadir web interface [291] was used to estimate the helix propensity of the fibrinogen coiled-coil region under physiological conditions. Only sequences that contained the full coiled-coil regions where included in this analysis. As a test case, the helix content of a poly-alanine peptide under physiological conditions was calculated and gave 100% helix content.

Another important feature in proteins are flexible and disordered regions. Such regions can be identified using the disEMBL program [292]. This program asses the probability that a given residue is in a disordered region based on a neural network trained on over 1000 reference PDB structures. This work focuses on two measures calculated by disEMBL. The first is the coil probability, i.e. the probability that a residue is not part of a DSSP secondary structure element. The second is the hot coil probability, i.e. the probability that a residue is part of a coil and likely mobile. The mobility is judged according to the temperature factor of the crystal structures used in the training set. Only

this last score corresponds to truly disordered regions. The results of this extensive comparison are presented in section 5.2.2.

4.2.9 *Denaturation and Reorientation at surfaces*

To quantify the behavior at the surface, the minimal distance between different regions (βC-, γC-domain, E-region) of the protein and the surface were measured. Futhermore the height of the center of the region over the surface was measured. If the minimal distance was below 0.25nm, the part of the protein was seen as contacting the surface. To define contacts of individual residues, a contact was counted as a heavy atom closer than 0.5nm to the surface (not including the counterion layer). A contact was seen as significant if it persistet for more than 1ns.

One of the defining effects of protein adsorption are strucutural changes. These changes are caused by residues rearranging to form the most favorable contact with the surface. In section 2.3.2, it has been discussed that such changes can be picked up by AFM experiments as an increase in diameter of the protein accompanied by a decrease in height. The detection of denaturation events is done in this work in a similar way. The denaturation is quantified by the difference in minimal distance between protein and surface and the center of the respective domain. If this distance decreases significantly it means that the region is denaturing. The difference to the first adsorbed conformation was recorded as a measure of change. Because the βC- and γC-domain as well as the E-region are approximately spherical, a simple rolling motion over the surface will not reduce the distance significantly.

Another important factor are reorientations at the surface. These are quantified by a height difference between two points in a region. As a consequence, the orientation is measured in units of length. For the E-region the center of the Fp-tethers (α27–47 and β57–80 in both chains) was chosen as the top most point and the center between the γ-chain connections (γ14–21) as the bottom point. The top of the D-region was defined as the middle between the centers of the βC- and γC-domain. The bottom was chosen as the disordered region connecting the globular domains to the coiled-coil region (α62–68,β193–202,γ136–144). Two more points were defined in the coiled region to monitor its orientation. The center of α88–99 and β109–142 was one point and the center of γ63–75 the other.

4.3 EFFECT OF COLLECTIVE VARIABLE CONSTRAINTS AND FINITE SIZE EFFECTS

As was mentioned in section 4.1.1 the rectangular shape of the simulation boxes makes it necessary to introduce constraints as otherwise a rotation of fibrinogen will lead to finite size effects. Such effects occur when two parts of the protein get close to

opposite box boundaries. In this case, the periodic boundary conditions lead to unphysical interactions between those parts. These artificial interactions disturb the dynamics of fibrinogen. The dominant finite size effect observed in the simulations is a collision between periodic images of fibrinogen. This can lead to deflection or aggregation of the periodic images depending on which parts of fibrinogen collide and is visually easy to spot in the trajectories.

For the production runs of protomer and dimer systems collective variable constraints were employed to keep the D-E and D-D axis respectively aligned to the long axis of the simulation box. These constraints were implemented using the colvar module in NAMD. This module calculates a collective variable, i.e. a quantity that is a function of the coordinates of several atoms, and applies a force determined from this variable. The force is not applied to any single atom but rather spread out over all atoms defining the collective variable.

Two different collective variables were used. Both compute the vector connecting the globular domains and bias it toward the long axis of the simulation box. The first constraint uses the projection of the connection vector on the plane perpendicular to the long axis of the simulation box (distanceXY option) and restricts it with a harmonic potential to the value found in the crystal structure. The force constant of the harmonic potential was chosen as $5\mathrm{kcal/mol}$. The second method constrains the direction of the vector to the long axis of the box (directionDir option). The force constant was chosen as $500\mathrm{kcal/mol}$ in this case, to reproduce forces of similar magnitude as for the distanceXY collective variable. These constraints only prevent a rotation of the protein around an axis perpendicular to the long axis of the simulation box but do not bias it towards the straight conformation. Both options decrease the simulation speed significantly.

By a comparison of simulations of the unglycosilated protomer using collective variable constraints with unconstrained simulations it was verified that no influence on the dynamics of the protein was present. This is demonstrated by comparing the energy contained in the collective variable to the fluctuations in the other energy terms in the simulation. The energy fluctuation were computed as deviation from the average energy of the protein over the trajectory. The energy contained in the colvar is on average $0.3\mathrm{kcal/mol}$ for the distanceXY with a standard deviation of $0.5\mathrm{kcal/mol}$. For the directionDir colvar the average energy is $3.8\mathrm{kcal/mol}$ with a standard deviation of $6.0\mathrm{kcal/mol}$. If the maximum values of the colvar energy are compared to the thermal fluctuations in the other energy terms of the affected regions, one sees that even the largest bias energies are smaller than the fluctuations for most of the simulation time (see Table 10).

This effect is even more dramatic if the biasing energy is compared to the absolute value of the different energy contributions.

	XY	*dir*
bond	*86.3%*	*78.9%*
angle	*91.6%*	*87.9%*
dihedral	*82.3%*	*75.6%*
improper	*60.2%*	*50.4%*
electrostatic	*96.9%*	*97.8%*
VdW	*86.7%*	*84.8%*

Table 10: Fraction of the trajectory in which the energy fluctuations in the parts of the protein on which the collective variable acts are larger than the maximal energy in the collective variable constraint. The distanceXY (XY) and directionDir (dir) constraint types are given.

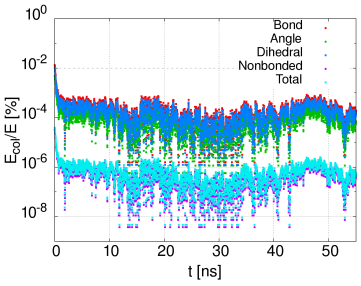

Figure 23: The collective variable energy contribution as a fraction of the other energy contributions in the system in a typical protomer trajectory.

Figure 23 shows that the collective variables contribute only a fraction of $10^{-5}\%$ of the other energy terms.

To exclude biasing effects arising from the application of the (weak) biasing forces over a long time the overlap of the PCA eigenspaces were calculated according to eq. (4.6). The PCA was performed on 40ns of simulations of unglycosilated fibrinogen without any collective variable constraints, and the first 40ns of trajectories with distanceXY (2 trajectories) and with directionDir (2 trajectories). The results of this calculation are given in Table 11.

	w/o	XY, 1	XY, 2	dir, 1	dir, 2	gXY, 1	gXY, 2
w/o	1						
XY, 1	0.365	1					
XY, 2	0.468	0.463	1				
dir, 1	0.259	0.224	0.225	1			
dir, 2	0.414	0.443	0.478	0.483	1		
gXY, 1	0.638	0.459	0.573	0.265	0.432	1	
gXY, 2	0.458	0.542	0.551	0.249	0.468	0.377	1

Table 11: Overlap of the spaces spanned by the 10 dominant PCA modes on the given trajectories of unglycosilated, unconstrained (w/o),distanceXY constraint (XY), directionDir constraint (dir) and glycosilated fibrinogen with distanceXY (gXY) constraint.

The overlap between unconstrained and constrained simulations is comparable to the overlap between runs with the same collective variable constraints. Therefore one can conclude that the dynamics are not influenced significantly by the collective variables.

4.3.1 *Avoiding finite size effects*

As was just seen, the long axis of fibrinogen can be held parallel to the long side of the box without influencing the dynamics by applying weak restraints on the center of mass of the E- and D- regions. However, the diameter of the protein structure can increase due to bending motions described in section 5.2 potentially leading to finite size effects. This is especially dramatic for dimer simulations.

One way to avoid finite size effects without increasing the simulation box too much would be to use a Parrinello-Rahman baro-

stat [293]. This allows the readjustment of the sides of the simulations box so that the long side could be decreased while the box could grow wider. This would only work for limited bending events as the initial volume is simply too small for large bending events independent of a reshaping of the box. Unfortunately, the Parrinello-Rahman barostat is not implemented in NAMD.

Alternatively, the size of the simulation box could be increased from the beginning. A dimer simulation box of sufficient size can be constructed under the assumption that the central region of fibrinogen remains aligned to the long axis of the simulation box. This can still be done using collective variable constraints. Such a box would have the dimensions of $49.3\text{nm} \times 18.9\text{nm} \times 22.7\text{nm}$ and contain 2,0062,289 particles. On 768 CPUs of the Mogon cluster (2.1GHz each) 1.61ns per day could be simulated. For the simulated dimer systems only 240 CPUs were used that achieved 2ns per day. This means the larger simulations box would require 4 times as many CPU hours as the box used. If a number comparable to the smaller systems (256CPUs) is used only 0.71ns can be simulated per day. This would mean simulations take nearly three times as long to reach the desired length. Considering that the current simulations took months to complete a larger simulation box was thus not feasible. This consideration does not include waiting times in the cluster scheduling system that would also increase with the increased CPU requirements.

An alternative program was developed that can resize the box without disturbing the close environment of the protein. This program works by extracting the protein and its solvation shell, defined as the solvent within a distance of 1.25 times the van der Waals cutoff (here 1.75nm), from the simulation box and inserting it into a new rectangular water box with adjusted dimensions. The new water box is equilibrated around the solvation shell and the simulation restarted. In practice, this did not provide a viable solution as the flexibility of fibrinogen necessitated frequent applications of the program so that the net simulation speed was even slower than the large box.

As no good solution for the problem was found, the simulation were performed with the box sizes and collective variable constraints discussed previously. The resulting trajectories were checked for periodic collisions. Collisions were specified as conformations in which the minimal distance between periodic images was less than 2.0nm. If such collisions were found, the simulations were stopped. The resulting trajectory lengths at which the simulations were stopped range from 14ns to 199ns and are given in Table 6 and Table 7. Stopping the simulations in this way might introduce a bias toward certain bending motions that do not lead to periodic collisions.

5

SOLUTION DYNAMICS

The first step in understanding the dynamics of fibrinogen is to identify structural motions with a physiological relevance. These movements likely occur as motions of subunits of the molecule that move as approximately rigid objects. Such subunits can be identified by atomistic simulations of fibrinogen in solution. All atom simulations in solution have the added benefit of providing a baseline behavior from which deviation can be observed close to a surface. It has already been noted that fibrinogen is a very flexible molecule [17, 19] so that structural fluctuations should be observable on the timescale accessible to atomistic MD simulations. Furthermore, simulations in solution can provide information on the importance of different parts of the protein for its dynamics. To reveal important features, four different subsystems have been investigated, namely the fibrinogen homodimer, with and without glycosilation and the protomer with and without glycosilation, i.e with and without the carbohydrate cluster attached to the β-pocket (see Fig. 7 in section 2.1.2). The glycosilation state was investigated because it has been speculated that it can affect the flexibility of fibrinogen [19]. While it will be shown that this is not the case, simulations of the different glycosilation states lead to the identification of a possible allosteric effect.

5.1 INFLUENCE OF DIFFERENT FACTORS ON THE DYNAMICS

To characterize the similarity in the correlated motions of fibrinogen, the PCA overlap defined in eq. (4.6) was calculated. For this to be a reliable measure, the simulations have to be sufficiently long to produce relevant motions. This was judged based on the cosine content of the dominant PCA modes that describe free diffusion in a free energy minimum [270, 294]. The signature of this behavior was not seen in the simulations used for the CPA calculations so that the simulation length was sufficient to extract modes describing significant motions of fibrinogen. Because the simulations are too short to sample the equilibrium distribution of fibrinogen conformations, it can not be excluded that some modes are missing.

The resulting PCA overlaps for individual protomer trajectories are shown in Table 11. It is clear that the presence of collective variable constraints does not seem to affect the PCA overlap

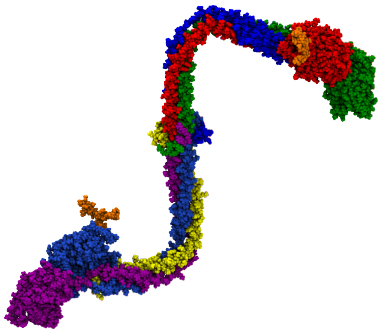

Figure 24: Snapshot from the simulation of the glycosilated fibrinogen dimer. A strong bending at two hinge points occurs during the simulation.

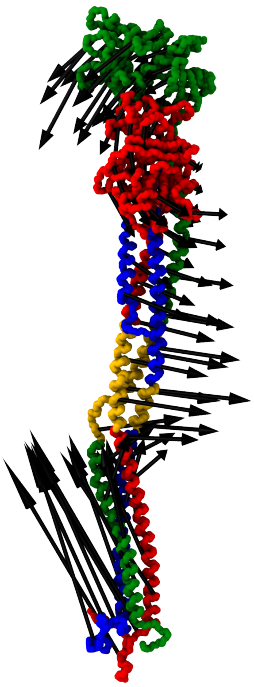

Figure 25: Visualization of the movement described by the dominant PCA mode in the fibrinogen protomer. A bending motion is described around the hinge region. The hinge identified by dyndom is highlighted in gold.

of the individual trajectories. To improve the sampling, the trajectories for the glycosilated protomer were merged, as were the trajectories of the unglycosilated monomer. In this case, the PCA overlap between the glycosilated and unglycosilated trajectories increased to 0.61. Consequently, it is concluded, that the presence of the carbohydrate clusters does not affect the large scale dynamics of Fg as observed on the timescale of the presented simulations. Therefore, trajectories can be analyzed together for large scale features. As will be seen later, the glycosilation does affect the dynamics on smaller scales.

Somewhat surprisingly, the dynamics of a free protomer are also similar to the dynamics of a protomer that is part of the fibrinogen dimer. In this case, the overlap in PCA space is 0.74. This means the motions of the two protomers in the dimer are very similar to those of free protomers and can be included in the analysis. The simulations of the dimer are, however, too short to say for certain if a correlation between the two protomers is present.

5.2 LARGE SCALE DYNAMICS: HINGE BENDING

The most obvious feature occurring in the simulation is a strong bending motion in both coiled-coil regions of fibrinogen. This is illustrated by the strongly bend conformation shown in Fig. 24.

5.2.1 *Characterization of the Hinge*

To characterize the large scale motion of fibrinogen, a PCA is performed. The dominant mode shows a bending at a hinge location depicted in Fig. 25. The PCA shows a clear bending mode and similar motions can be seen for the second and third mode as well. Higher modes include motions in the D-region as well as a torsion of the coiled-coil region. The projection of the PCA modes along the trajectory can be calculated along the simulation trajectories. For the longest protomer trajectories this projection is shown in Fig. 26. It can be seen that for the longest trajectories bending and unbending motions are observed. It is thus not the case that the fibrinogen molecule has a defined bend shape in solution. Rather the molecules fluctuates between bend and straight conformations.

To identify the residues responsible for the hinge, an analysis with dyndom is performed using the extremal structures of the dominant protomer mode. In this analysis, a well defined mechanical hinge point is identified (as opposed to a delocalized bending). The hinge point is given by residues α99–110, β130–155, and γ70–100. The hinge thus coincides with the partially unfolded region of the γ-chain and it can be inferred that the presence of the two proline residues γ70 and γ76 is the underlying cause for the flexibility of fibrinogen (see the discussion of α-helical structures in section 2.1.1). The proline residues break

up the α-helix in the γ-chain and thus induce a flexible defect in the stiff coiled-coil region. This flexible region can be on the compressed or the stretched side of the bend region as is shown in Fig. 27.

This flexibility is not a novel idea and has been suggested nearly 40 year ago [25]. In the X-ray crystal structure, the hinge region has been identified as residues α99–107, β126–131, and γ69–77. This region was identified on a fixed linear structure from variations in the four different protomers observed in the crystal unit cell as well as comparisons with other crystal structures [19]. Presumably the fact that only small deviations from the linear structure are observed in the crystal lead to the narrower definition of the hinge compared with the dyndom results. The extremal structures used in dyndom showed strong deformations in α- and γ-chains that were not observed in the crystal structures. Outside of the hinge the coiled-coil region behaves as a rigid object without further large scale deformations.

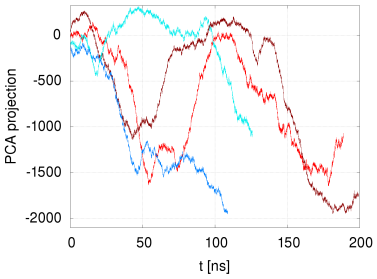

Figure 26: Projection of the dominant PCA mode along the two longest trajectories of the glycosilated (red) and unglycosilated (blue) protomer.

Beside the bending motion of the hinge, a rotation of the outer regions around the central region is observed. The hinge serves as the pivot point for this rotation. The hinge behavior can thus be characterized by two angles: a bending angle γ and a dihedral angle φ. The bending angle is defined as the angle between the centers of the residue groups (α50–58,β82–90,γ23–31), (α99–110, β130–155,γ70–100) and (β200–458,γ140–394), i.e. a point in the E-region, the hinge region and the D-region. The dihedral angle is defined by splitting the group in the E-region into two regions (α50–58,γ23–31) and (β82–90,γ23–31).

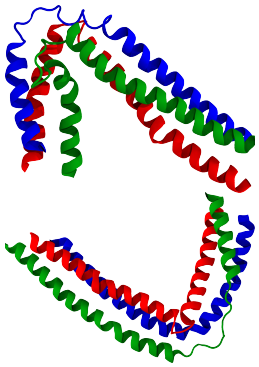

Figure 27: Two conformations of the hinge region where the flexible region is either compressed (top) or stretched (bottom).

The angles γ and φ are a good characterization of the hinge motion as judged by their high correlation with the PCA modes. The linear correlation coefficient of the bending angle γ with the first PCA mode of the protomer is 0.82 and -0.51 for the second. The dihedral angle φ has a correlation of 0.51 with the first mode and 0.66 with the second mode. In the dimer case, the correlation is also good. The correlation coefficient between φ and the first PCA mode is -0.70, while γ has a correlation coefficient of -0.94 with the second PCA mode. Because the rotation and bending motions are nonlinear, a perfect correlation was not expected instead the motion has to be expressed as a linear combination of PCA modes. This is supported by the fact that the first three PCA modes span a 2D hemispherical surface and can thus be described by two angle parameters.

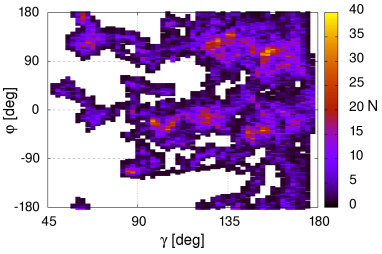

Figure 28: Histogram of hinge bending (γ) and dihedral (φ) angles.

Both angles can be measured along the trajectories and a histogram of the occurring values is shown in Fig. 28. The histogram shows that two easy bending direction exist that differ by about 180° in φ. This means the strongest hinge bending events happen in a plane. It should be noted that angles above $\gamma = 120°$ have a fairly broad distribution so that in this case the outer regions can rotate. Care should be taken before putting too much faith into the detailed features of this landscape. As was mentioned earlier, the sampling of the hinge conformations is far

from exhaustive in the atomistic simulations so that it does not necessarily represent the underlying free energy landscape correctly. One feature that is almost certainly due to the limited sampling is the high number of conformations in the vicinity of the starting structure at $(\gamma, \varphi) = (150°, 120°)$.

Another limitation is that neither the Aα-regions nor the γ-chain carbohydrate clusters are present in the simulations. Both of them are close enough to the hinge to interact and limit the bending or dihedral motion. The flexibility of the Aα-chain makes such a targeted interaction unlikely and the second carbohydrate cluster is located in the N-terminal direction from the hinge while the bending modes show deformations in the C-terminal direction. The missing parts of the protein might thus limit the maximal bending motion but it seems unlikely that they could stabilized the coiled-coil region enough to form a rigid structure. Countering these stabilizing effect is the fact that the used version of the CHARMM force field is known to stabilize helical conformation [155]. Larger unfolding events resulting in a greater destabilization of the coiled-coil region are thus not implausible in real systems. In addition to these static characteristics, the simulations also allow a glimpse into the dynamics.

The bending time was estimated by the time span it took the protein to first reach a conformation with $\gamma < 120°$ from a conformation with $\gamma = 160°$. This value was chosen as it corresponds to the starting structure. Measuring from this point instead of the start of the simulation eliminates waiting times that sometimes occur. Waiting times occur in 8 out of 13 cases and range from 10 to 143ns (median 11ns). The resulting bending time was averaged over all trajectories and gives a bending time of (18 ± 3)ns. The bending is thus very fast for the movement of such large regions.

The presence of the hinge immediately calls the model in Fig. 12 for the interpretation of adsorbed fibrinogen conformations into question. The dimer simulations also do not show a bending at the E-region that is assumed in this model. Despite the hints for flexible hinges identified early in fibrinogen structure investigation, a model with deformations occurring in the E-region has been used in the interpretation of many adsorption experiments. This model should be improved to take the two symmetric hinges in the dimer into account. How this can be done is discussed in Chapter 8.

5.2.2 *Conservation of the Hinge and Helicity Patterns across Species*

Conservation

The significant effect of the hinge on the behavior of fibrinogen described above hints at an important physiological function. As such, this feature should be highly conserved during the evolution of blood clotting. To investigate the evolutionary conservation of the hinge, fibrinogen sequences of all major branches

of vertebrates as identified in section 4.2.8 were considered. An analysis of the phylogenetic tree created based on the alignment of the γ-chains shows that it is in line with the known evolutionary relationship between the species (see Fig. 97 in Appendix A).

In all species, except for the lamprey, the hinge region identified in human fibrinogen corresponds to a proline rich region. Most of the tetrapod species have more than one proline residue in this region while most of the fish in the sample have only a single proline. The difference in occurring prolines thus identified corresponds to three major stages in the evolution of the clotting cascade in which the different pathways were invented (appearance of intrinsic pathway in jawed vertebrates, addition of contact pathway in tetrapods; see section 2.2.1). A representative sample of aligned hinge sequences is shown in Table 12. For tetrapods the extra prolines are inserted in the five residue gap between the two prolines occurring in human fibrinogen. The exact hinge sequence is not well conserved, suggesting that the function of this region is to create a structural change by breaking the α-helix, rather than forming an interaction site for another protein.

A very surprising finding is the unusual hinge sequence for cetaceans, which have lost their first proline and added two glycine residues flanking the remaining two prolines. The shorter separation of the prolines and added flexibility provided by the glycines would suggest a one residue extrusion from the coiled-coil region (see the discussion of coiled-coil structures in section 2.1.1) and thus a very small hinge if it is present at all. Similarly, the hinge region for the bat *M. davidii* contains only a single proline. What implications this has for the hinge and clotting behavior or clot lysis is unclear.

Human	...TYNPDESSKPNMIDAAT...
Chimpanzee	...TYNPDESSKPNMIDAAT...
Mouse	...YYNPDQPPKPGMIDSAT...
Cow	...SYNPDQPSKPNNIESAT...
Dusky dolphin	...SYHSDGPAKPNGIESAT...
Harbor porpoise	...SYHSDGPAXPSGIESAT...
Sperm whale	...SYRSDGPAKPSGIESAT...
Horse	...SYNPDEPPKTGRIDAAT...
Dog	...TYNPDEPPKPNRVVGAT...
David's myotis	...FYSTDGLPKSDRIEASI...
Fruit bat	...NYDPNEPPKPNKIDSAT...
Tasmanian devil	...GQGSD-PRPQNVINDAM...
Platypus	...G-GPDRSREPQNRQGVT...
Duck	...LYPPDKQQQQQSVEGFT...
Collared flycatcher	...LYPSDKQTLPNSIDDFT...
Softshell turtle	...LSPPEK-TQPSTIDGFT...
African clawed frog	...SGKKPATSPQTAIDPMT...
Chinese Sturgeon	...SQTQA--QKSSP-DLYI...
Ayu	...SVTSA--QKSSSPDQYF...
Ghostshark	...SNVVIK---KQLPDTFI...
Sea lamprey	...DVRITRDEAQIIKDSGQ...

Table 12: Amino acid sequence of the fibrinogen γ-chain in the hinge region for selected species. A complete list can be found in Table 27 and Table 28 in Appendix A. Prolines and glycines are highlighted for clarity.

In a second step, the α- and β-chains have been aligned. It is noted that the conservation of the N-terminal disulfide ring is stronger than the for the C-terminal ring. The sequence of the first one being identical in nearly all species while the latter is more variable especially when compared between tetrapods and fish. This is in line with the knowledge that the N-terminal ring is more important for dimer formation while the C-terminal ring is not necessary to form the dimer[1].

Other proline residues occurring in the coiled-coil region are β92 and β162. The later is highly conserved, being present in all sequences, except for the lamprey, while the former is only well conserved in placental mammals. The β92 proline is thus a signature in the most recent evolutionary change in the clotting cascade. Only half of the other species show proline early in the β-chain but the position varies and the occurrence seems to be independent of species relationships. The position of the β162 proline coincides with the end of the helical region of the

1 It is however necessary for excretion of the protein from mammalian cells [31].

C-terminal α-chain that is folding back onto the coiled-coil region.

In birds, but not reptiles, an additional proline is found at the position equivalent to $\alpha 96$, i.e. in the hinge region. Some birds, reptiles and bony fish show an additional proline at the $\gamma 34$ position.

The occurrence of novel proline residues thus seems linked to the major stages in the evolution of the clotting cascade. This supports the assumption that the flexibility induced by the introduction of prolines has physiological relevance for blood clotting.

Helicity

To judge the effect of different sequences on the existence of a hinge, the helicity of the sequences has been estimated using agadir.

The predicted helix propensity for human fibrinogen is shown in Fig. 29, while the results for all species are shown in Fig. 30 and Appendix A.2. The predicted helix content clearly shows that the coiled-coil region is made up of alternating helix friendly and unfriendly regions. A pattern that is evolutionarily well conserved. Such helix friendly regions are known as trigger sequences and ensure the correct folding and association of coiled-coils [15]. These trigger regions form nucleation sites for the coil to helix transition and provide strong interactions for oligomerization. Moreover the helix unfriendly regions 15, 36 and 59 residues after the disulfide ring coincide in all three chains. The last of these gaps corresponds to the previously identified hinge region. The gap 36 residues after the disulfide ring is the only one not associated with a proline, while no clear gap is associated with proline $\beta 162$.

To identify the helix unfriendly regions more clearly, the agadir score for all three chains were combined. On average the gaps 15, 36 and 59 residues after the disulfide ring are 4.6, 3.4 and 5.3 residues wide and have a combined helix content of 0.89%, 1% and 0.85%. Only in pig, naked mole rat, myotis and lamprey fibrinogen is the gap 36 residues after the disulfide ring more pronounced than the gap 59 residues after the disulfide ring. Considering that the gap associated with the hinge region is deeper and wider than the other gaps, this supports the hypothesis of a single hinge. Furthermore this gap has more prolines in the γ-chain that can disrupt any templating effects from the surrounding regions. It has to be mentioned that positions of prolines in other species in the gap 15 residues after the disulfide ring could indicate a somewhat weaker hinge close to the E-region. However, a single proline would be unlikely to create an extended and well defined disordered region. Alternatively this region could serve as a seed for a guided helix-coil transition under deformation of the molecule in the fibrin fiber. This can be motivated by Levines helix-coil transition model (see sec-

tion 3.3.2) [218–220]. In Levines model the presence of prolines or regions with low helix affinity would correspond to the introduction of impurities in the Ising field. These impurities form nucleation centers for the helix-coil transition. It is known that such impurities lead to domain wall pinning in the Ising model [295] and thus fix the point at which the helix-coil transition occurs. As a consequence, the nucleation of the coiled phase and with it the bending of the coiled-coil structure would not happen at a random position but rather at a predefined point.

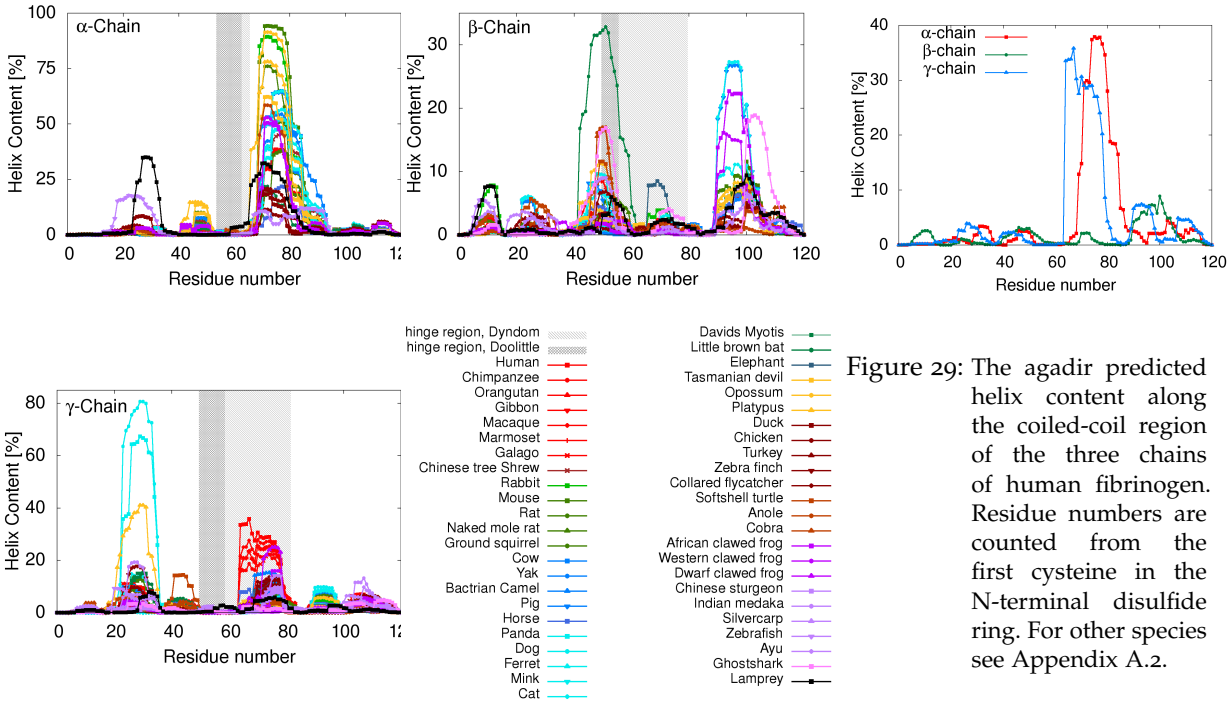

Figure 29: The agadir predicted helix content along the coiled-coil region of the three chains of human fibrinogen. Residue numbers are counted from the first cysteine in the N-terminal disulfide ring. For other species see Appendix A.2.

Figure 30: The agadir predicted helix content along the three chains of the fibrinogen coiled-coil region. Residue numbers are counted from the first cysteine in the N-terminal disulfide ring. The shaded regions correspond to the hinge identified by dyndom in section 5.2 and in [25].

Support for this hypothesis comes from steered molecular dynamics simulations of forced fibrinogen unfolding [249, Supporting Information]. During pulling, the coiled-coil region does not unravel evenly from the hinge outward. Rather a piece of the coiled-coil region, close to the E-region, starts unfolding before all of the intermediate region has lost its helical structure. No such separate unfolding event seems to be associated with the gap located 36 residues after the disulfide ring or the proline at β162. These SMD simulations also show that the α-chain loses its helical conformation in the hinge region before the β-chain does. This is consistent with the observation that the β-chain has a somewhat higher helix propensity in this region and Fig. 27 that shows a larger deformation in the α- than in the β-chain.

It can also be noted that in fish the helix propensity is much more evenly distributed along the helix friendly regions in the coiled-coil region, whereas tetrapods usually have a dominant peak after the hinge.

A complementary way to asses the flexibility of fibrinogens coiled-coil region is to characterize potentially disordered parts using disEMBL [292]. Figure 31 shows the probability to have coil and hot coil regions along the coiled-coil region of human fibrinogen. While the coil score identifies the same regions as helix unfriendly as the agadir analysis, the hot coil score clearly identifies the hinge as the only high mobility region.

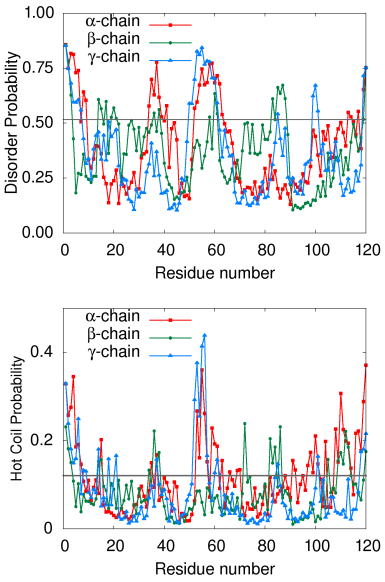

Figure 31: The disEMBL predicted probability for coil (top) and hot coil (bottom) regions along the coiled-coil region of the three chains of human fibrinogen. Residue numbers are counted from the first cysteine in the N-terminal disulfide ring. The gray line is the threshold below which no disordered region is expected. For other species see Appendix A.3 and A.4.

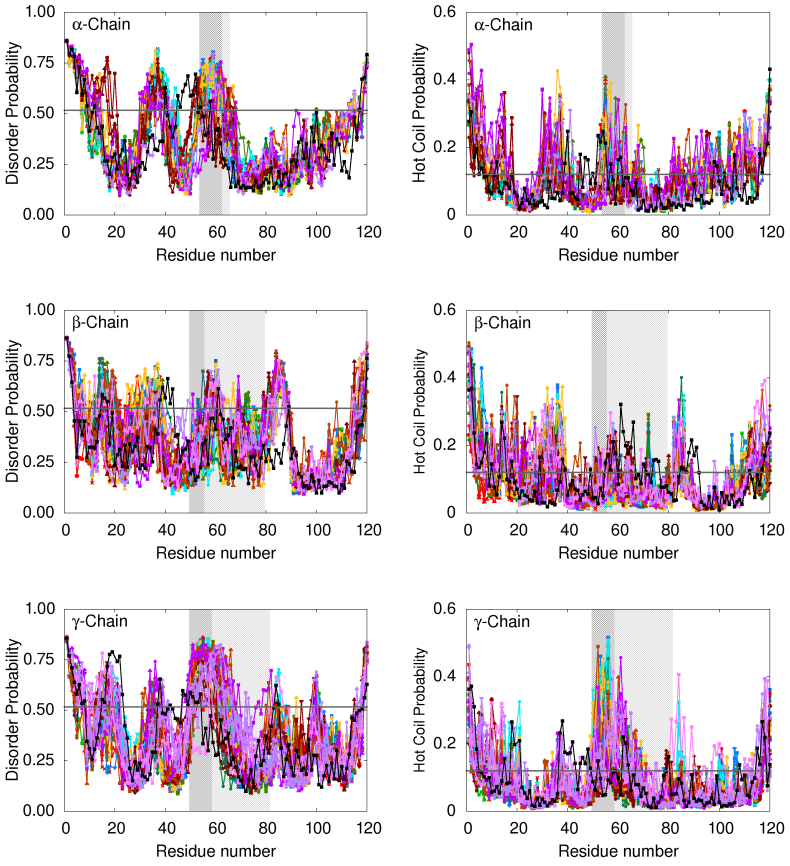

Figure 32: The probability to find the three fibrinogen chains in a coil region (left) and a hot coil region (right). For a legend see Fig. 30. Residue numbers are counted from the first cysteine in the N-terminal disulfide ring. The gray line is the threshold below which no disordered region is expected.

A comparison of the disEMBL results for different species is given in Fig. 32. The disEMBL score is more consistent between species than the agadir prediction. Again, a clear alternation of regions with high and low disorder probability is found that is evolutionarily well conserved. The same gaps 15, 36 and 59 residues after the disulfide ring can be identified as likely disordered regions. Over all species the hinge region still has the

highest hot coil score although the region 36 and 85 residues after the disulfide hinge also have significant scores.

More interesting is the comparison for species for which the presence of a hinge was more ambiguous from sequence data and agadir results, such as bats and lampreys. There is no significant difference between the disEMBL scores for bats and other mammals and the hinge region is dominant for the hot coils. For lampreys a single extended hot coil region is found that ranges from 34 to 66 residues after the disulfide ring. This region is less pronounced than for other vertebrates.

The disEMBL algorithm can also be applied to the γ-chain fragments of cetaceans. In this case, a hot loop region at the end of the fragment, corresponding to the hinge region, is identified. The overall disorder pattern being in line with other mammals. Thus, it seems likely that all jawed vertebrates show the same hinge behavior. It remains an open question if lamprey fibrinogen also has a well defined hinge.

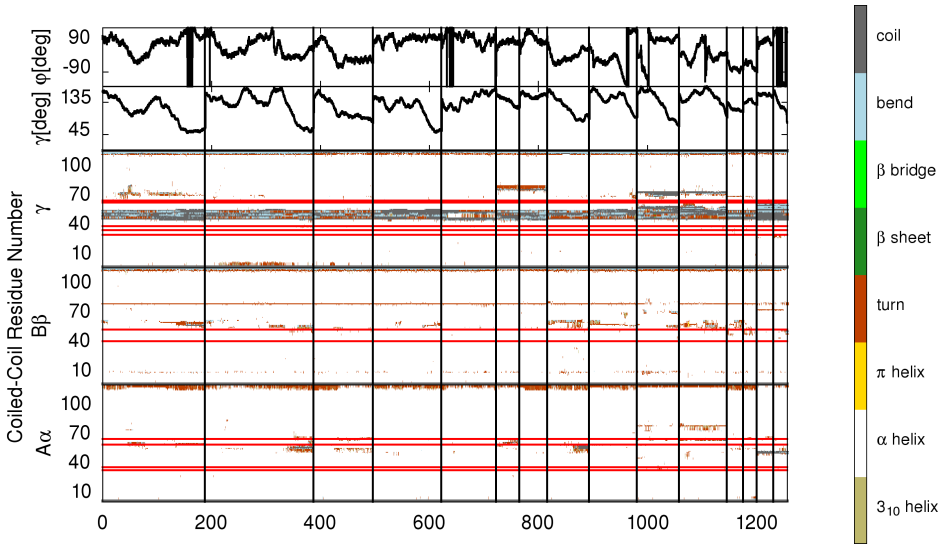

Figure 33: DSSP secondary structure assignment in the coiled-coil region of fibrinogen along the trajectory. The unfolded regions around residue number 80 in the γ-chain comes from the second protomer of the dimer simulation. The red lines show the position of plasmin cleavage sites, the black lines show the start of a new trajectory. The trajectories are shown in the following order: glycosilated protomer, unglycosilated protomer, glycosilated dimer, unglycosilated dimer.

To support the agadir and disEMBL calculations, the secondary structure of the coiled-coil region can be monitored along the simulated trajectories. The secondary structure is assigned using the DSSP algorithm and the results are shown in Fig 33.

The assigned secondary structure shows that the α-helix is disrupted most in the hinge region. In general, only weak disturbances of the helical structure of the α- and β-chain occur in our simulation. The effect a proline can have in breaking helical templating effects is illustrated by the fact that the helical structure is more disturbed in the region 80 residues after the disulfide ring

than in the region 36 residues after it, despite the former being helix friendlier. Similarly, the helical structure around residue number 15 is more disturbed in the β chain than it is in the α- or γ-chains, where no prolines are present at this point.

5.2.3 *Disease Mutations in the Hinge Region of Human Fibrinogen*

The large motion of the hinge as well as its evolutionary conservation support a biological function of the hinge. If this function is disrupted by a mutation, it should lead to physiological problems.

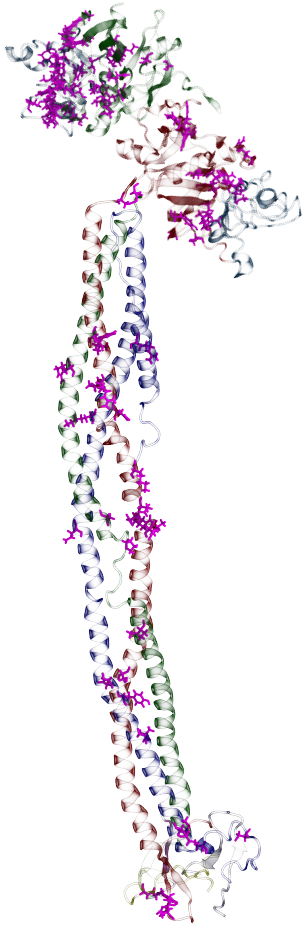

Figure 34: Known mutation sites of fibrinogen that lead to disease (pink) [296]. Many of the mutations in the D-region are located in the vicinity of Fibrinopeptide binding site (light blue).

The fibrinogen mutations database [296] contains many mutations that have been identified in human fibrinogen over the last 15–20 years. In this time span the availability of genetic sequencing techniques increased and allowed a more widespread application in clinical tests. Most of the mutations are identified in clinical cases were patients are screened for bleeding disorders during preparation for surgeries. Mutations act on the genes coding for the different peptide chains in fibrinogen and can take different forms. Substitution of a base pair can change the amino acid for which the sequence codes or change it to a stop codon that ends the DNA to protein translation. In the same way, a stop codon might be transformed into a sequence coding for an amino acid so that splicing sites for the gene might be disrupted [297]. The deletion of a base pair will lead to a shift in reading frame, that is the window identifying the three base pairs for an amino acid gets shifted with respect to the sequence (see Fig. 1 in section 2.1.1), leading to a translation of gibberish.

Fibrinogen mutations that do not interrupt protein translation (i.e. mutations to a stop codon) are shown in Fig. 34. As can be seen, many mutations are associated with the ligand binding sites. This does not necessarily indicate a higher mutation rate for these residues. Instead, it is more likely that mutations in other parts of the protein are not identified because they do not disrupt fibrinogen function. Therefore, no bleeding disorder occurs and there is no reason to look at the fibrinogen genes. While many mutations are known, the bleeding disorders associated with fibrinogen mutations are rare conditions affecting less than 1 in 90,000 patients. Not all mutations are located in the D-region. Some have been identified in the coiled-coil region and especially in the hinge region. These mutations will be described in the following.

Only one mutation of the α-chain is known to occur in the hinge region. This mutation is a substitution from aspartic acid to asparagine at position α106. This mutation leads to impaired fibrin formation, thinner fibrin fibers, longer clotting and lysis times and reduced platelet aggregation. It has been speculated that this mutation disturbs the flexible regions to cause these effects [47].

The β-chain has three known mutations in the hinge region. The first is a deletion of five residues from β137 to β141. This deletion leads to a decrease in fibrinogen functionality resulting in longer clotting times and faster lysis [49]. The second mutation is a substitution of lysine with asparagine at position β148 that has not been associated with any physiological problems [298]. The last mutation occurs at β160 and gives rise to an extra glycosilation site. The additional carbohydrate cluster attached at this position leads to a changes in fibrin structure [299].

In the γ-chain, two mutation are associated with the hinge region. A shift in the DNA reading frame is known to occur at the coding sequence for the proline at position γ76. This reading frame shift deletes large portions of the γ-chain between γ76 and the C-terminus. Such a major mutation makes it unlikely that the γ-chain can correctly associate with the other two chains to form a functional fibrinogen molecule [300, 301]. Another mutation in the γ-chain occurs at position γ82 where an alanine residue is substituted with glycine. While the severity of this comparatively common mutation varies [302–304], it seems to be the case that fibrin formation proceeds unhampered. The problem occurs due to an increased degradation rate of fibrinogen which leads to hypofibrinogenemina [305].

To estimate the significance of these mutations, the number of mutations in the hinge region are compared to the surrounding coiled-coil region. For this comparison only mutations have been considered that do not delete significant parts of the protein (e.g. [297, 300]) or add glycosilation sites (e.g. [299, 306]) as these major events affect much more than the hinge. Under these conditions, eight mutations are identified in the hinge region.

Two mutations early in the α-chain disrupt the correct formation of fibrinogen dimers [307, 308] and are also not considered. Even without these exclusions, seven mutations where identified in the coiled-coil region surrounding the hinge. Two of those are located in the α-chain. One of them is a deletion of an asparagine residue at position α80 and leads to unstable and porous clots [309] while the other is a leucine to histidine substitution that makes clots more resistent to lysis [310]. The β-chain contains another deletion at β111 causing augmented lateral aggregation [311] as well as three substitutions at positions β118 (methionine to lysine), β165 (leucine to proline) and β166 (arginine to cysteine). The first two lead to hypofibrinogenemia [312] while the last leads to the formation of disulfide bonds between different fibrinogen molecules [313, 314]. The last mutation is located in the γ-chain at position γ114. This is a tyrosine to histidine mutation and leads to impaired lysis [310].

While the hinge region makes up 18% (66 of 360 residues) of the residues in the coiled-coil region, it contains more than half of the residue mutations in the this region. This ratio of mutations seems to support a functional role for the hinge in fibrin formation. But it should not be forgotten that plasmin cleavage

sites are located close to the hinge [25, 37]. These sites could also be affected by the mutations and cause physiological problems, especially with clot lysis. Indeed, many of the mutations in the hinge region lead to problems with lysis or fibrin formation.

5.2.4 *Experimental Evidence for the Existence of the Hinge*

Until now, mostly circumstantial evidence for the existence of the hinge has been presented. It was already mentioned that crystal structures provide evidence for limited flexibility of the hinge in the case of human fibrinogen [19]. The hinge has also been noted in bovine [315] and chicken fibrinogen [316]. In the last case, the hinge bending is very modest but the region itself shows a high mobility as judged by the temperature factors of the atoms.

Many other experiments do not allow an easy inference regarding the existence of a hinge. Only in the highest resolution imaging studies of adsorbed fibrinogen has it been possible to resolve structures between the globular regions. It can not be distinguished if these structures are the coiled-coil or the αC-regions. In cases where these structures are resolved, a kink in the middle of the connector region can be seen [17, 137]. This matches the presence of two hinges much better than a single hinge at the center or a continuous bending of the coiled-coil region. The highest resoultion fibrinogen structures to date have recently been reported [64]. In this study, fibrinogen was adsorbed on mica covered with a glycine-carbohydrate molecule to obtain a hydrophilic surface. The resulting structures have a resolution high enough to resolve the αC-regions. These images of fibrinogen also show clear kinks in the coiled-coil connector.

Another study that lends tentative support to the existence of a hinge has recently been published [37]. In this study, the exchange of protein hydrogen atoms with a deuterated solvent was investigated. Parts that show a fast exchange are assumed to be mobile and in good contact with the solvent. The identified hinge region shows such an accelerated deuterium exchange but so do other segments the coiled-coil region (e.g. β100–110 or α80-85). In Fig. 35 the parts of the coiled-coil region are highlighted that loose the helical secondary structure to some extend in the simulation and thus expose the backbone to the solvent. A comparison with the deuterium exchange profile shows that some regions that show a high exchange rate coincide with the deformed α-helical structure observed in the simulations while others do not (see Table 13).

In summary, it can be said that no conclusive measurement of the hinge has yet been performed. However, it should be noted that almost all experiments are consistent with the proposed hinge mechanism and no experiments contradict it. Considering how hard an unambiguous demonstration of the existence and flexibility of the hinge is because of the size of fibrinogen it is not

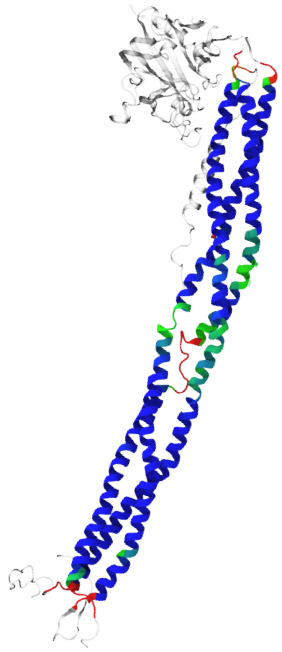

Figure 35: Average helix content in the coiled-coil region of fibrinogen averaged over all simulation trajectories. Colorscale: red=0, green=0.85, blue=1.

Region	*Simulation*	*Deuterium exchange*
α68–90	*no*	*yes*
α90–105	*yes*	*yes*
α105–125	*yes*	*no*
β85–87	*no*	*yes*
β89–91	*yes*	*yes*
β130–144	*no*	*yes*
β159–161	*yes*	*no*
γ30–35	*no*	*yes*
γ68–77	*yes*	*yes*
γ78–87	*yes*	*no*
γ93–105	*yes*	*yes*

Table 13: Comparison of regions where the helical structure is broken in the simulation and regions showing high deuterium exchange [37].

surprising that no explicit measurement has yet been performed. Proteins of this size can not be expressed in standard bacterial cell cultures which complicates mutational studies. Such studies could, e.g. replace the proline residues responsible for the hinge with alanine. If these mutations lead to a correctly assembled fibrinogen dimer, they should stabilize the coiled-coil region and possibly remove the hinge entirely. An effect should be clearly visible in measurements of the hydrodynamic radius of fibrinogen or the adsorbed structures at surfaces. An alternative method would be to perform FRET experiments by attaching dyes to the E- and D-regions. The problem with this approach is that dyes are usually attached to proteins with disulfide bonds but no free cysteine residues are present in fibrinogen. Additionally, most FRET dyes do not interact over large enough distances to get a signal even in strongly bend conformations.

5.2.5 *Possible Physiological Functions of the Hinge*

The evolutionary patterns and the disease mutations in humans suggest a physiological relevance of the hinge. One explanation might simply be that the hinge is a structural feature to provide flexibility. In section 2.1.3, it was already mentioned that the fibrin fibers branch of each other and form a space filling network. To do this, fibers need to branch of each other which requires some flexibility in the substructures. Whether this would require the individual molecules to deform, is not clear. An outer region of fibrinogen sticking out from the fiber due to the hinge bending would certainly provide a natural branching point. Due to the large systems sizes necessary to investigate this, this hypothesis can not be tested in atomistic simulations. The flexibility of fibrinogen would also be essential to accommodate it in the fibrin fiber structure where molecules are compressed in the core and stretched at the perimeter [58]. The effect of mutations on fibrin formation offers some support for this. However, the importance of the hinge for these effects can be questioned as it is known that fibrinogen molecules straighten during the assembly into protofibrils [63]. It has been speculated that the straightening is a consequence of the removal of the fibrinopeptides [17] but this seems at odds with the simulation. The fibrinopeptides were not present in the simulation and fibrinogen still bend.

Due to the many interactions of fibrinogen with other molecules described in section 2.1.2 it is plausible that the flexibility could impact the interaction of fibrinogen with other proteins. As mutations in the hinge also affect clot lysis, the hinge might be important for the interaction between fibrinogen and plasmin. It is known that the region around the hinge is the first target of proteases during the lysing of blood clots [19, 25] and in Table 1 in section 2.1.2 it can be seen that several plasmin attack sites are located close to the hinge. In chicken fibrinogen, the hinge even overlaps with plasmin attack sites [316]. The bending of

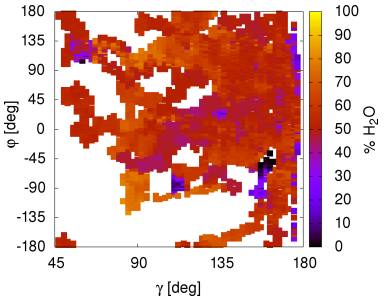

Figure 36: Fraction of water in the environment of the plasmin attack site γ62–63 depending on the hinge conformation.

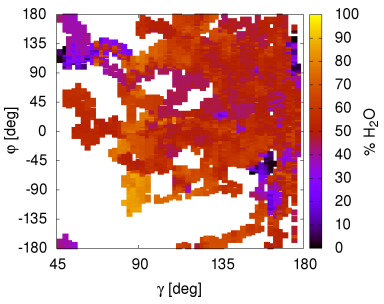

Figure 37: Fraction of water in the environment of the plasmin attack site β133–134 depending on the hinge conformation.

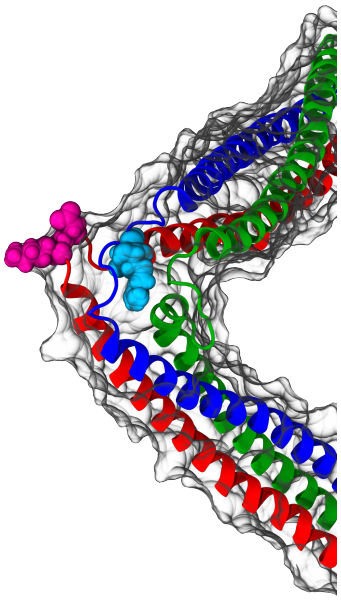

Figure 38: Bend conformation of the hinge that disrupts the secondary structure around the plasmin binding sites α104–105 (cyan) and β133–134 (magenta). The protein surface is shown in gray.

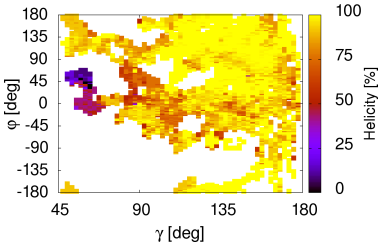

Figure 39: Averaged helicity of the plasmin cleavage sites α102 – 107 and β131–136 depending on the hinge conformation.

the hinge could then regulate the accessibility of these binding sites. To judge if this is the case, the surrounding of the plasmin binding sites was analyzed to compare the fraction of water to the fraction of heavy protein atoms in a 1nm sphere around the binding site. Because plasmin binds to a four residue sequence, two residues on the N- and C-terminal side of the plasmin attack sites were included in the definition of the plasmin binding site. Most of the binding sites show little change in their surrounding during the bending process and a typical example is shown in Fig. 36. However, the binding site at residues β133–134 shows a significant difference in accessibility in the strongest bend conformations of the two easy bending directions. This is shown in Fig. 37.

While the bending does not affect the accessibility of most plasmin attack sites, it can disrupt the helical secondary structure of the sites and thus expose the backbone for cleavage by plasmin. As can be seen in Fig. 40 and Fig. 33, there is only a weak correlation between the hinge bending angle and the helix content of the plasmin attack sites. For none of the sites does the correlation between γ and the helix content exceed 0.4 and only the five sites closest to the hinge show correlation coefficients exceeding 0.15. The highest correlation is observed for the plasmin attack sites α104–105 (0.26 correlation coefficient) and β133–134 (0.37 correlation coefficient). The imperfect correlation can in part be explained by the dihedral motion of the hinge. If the plasmin attack sites are on the inside of the bending region, the secondary structure is less affected than in cases where they are on the outside. This is confirmed for the two plasmin binding sites showing the highest correlation with γ. In Fig. 39 it can be seen that the same minimum that increases the accessibility of the plasmin binding site β133–134 also significantly disrupts the helical structure of this binding site. One hinge bending event that shows such a bend conformation is shown in Fig. 38. In this case, the binding site on the β-chain is particularly exposed while the one on the α-chain is still buried.

The presence of an integrin binding site close to the hinge (see Table 2) [44] might also indicate an importance for platelet binding. The accessibility of this site is similar to Fig. 36. The same holds for the helicity of the site. The bending thus shows no effect on the plasmin binding site.

Many placental mammals have more prolines in the hinge region, often separated by two other residues. Such short proline rich motives are known to mediate interactions with cytoskeletal and signaling proteins [16] so that the hinge might have additional functions in these species.

Despite the several plausible mechanisms presented here, it is not possible to draw any definite conclusions regarding the physiological function of the hinge from the simulations. The combined evidence for an increased accessibility and disruption of

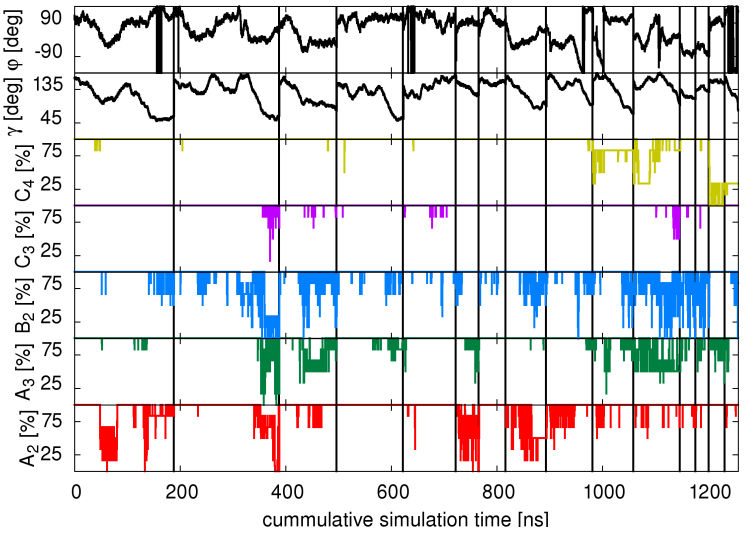

Figure 40: Fraction of the residues belonging to the plasmin attack sites that show a correlation greater than 0.1 with the hinge bending angle γ. Vertical lines denote the start of different simulations in the same order as in Fig. 33. The binding sites are labeled as follows: $A_2 = \alpha 102 - 107$, $A_3 = \alpha 108 - 113$, $B_2 = \beta 131–136$, $C_3 = \gamma 60–65$ and $C_4 = \gamma 83–88$. The disruptions in C_4 correspond almost entirely to the larger unfolded region in the second protomer of the dimer simulations.

secondary structure for some plasmin binding sites favors an explanation where hinge bending is an important factor in fibrinolysis.

5.3 SMALL SCALE DYNAMICS: FP BINDING POCKETS

The hinge bending investigated so far is a very large scale phenomenon for atomistic simulations. Nevertheless, the simulations also give information on smaller scales. The smallest scale features of interest for the aggregation of fibrinogen are the ligand binding pockets in the D-region (see Fig. 6 in section 2.1.2). Their dynamics are investigated here.

5.3.1 *Dynamics of the β-pocket*

As the pocket is responsible for binding a peptide ligand (see section 2.1.3), it is important to identify flexible parts that can deform to stabilize a ligand. Another important point is the movement of the residues responsible for ligand binding. These residues have been identified and a crystal structure with a bound ligand is available in the PDB (PDBID 1FZC) [24, 29].

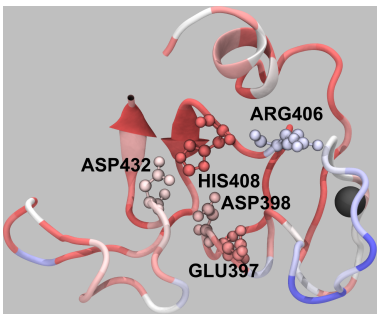

Figure 41: RMSF of the residues in the glycosilated β-pocket. The values range from 0.60Å (red) to 3.87Å (blue). The calcium ion is shown in its crystal position.

Flexible Regions and Modes of Motion

The flexible parts of the pocket are identified by the RMSF depicted in Fig. 41. The overall RMSF pattern is similar in glyco-

silated and unglycosilated simulations and identifies three large moveable parts. Two are loop regions one of which is part of the calcium binding site (β383–393, CA-loop) while the other sits on top of the two α-helical strands of the α-chain (β425–433,CC-loop). The third flexible part is the α-helix to which the glycans are attached. A fourth small loop (β395–399) is not flexible in the glycosilated simulations. These regions match those identified in previous simulations of the fibrinogen D-region [252].

To characterize the motions of these flexible parts, a PCA is performed. The dominant two PCA modes are shown in Fig. 42. Both results are similar in the glycosilated and unglycosilated simulations as indicated by a PCA overlap of 0.76 for the subspace containing 95% of fluctuations.

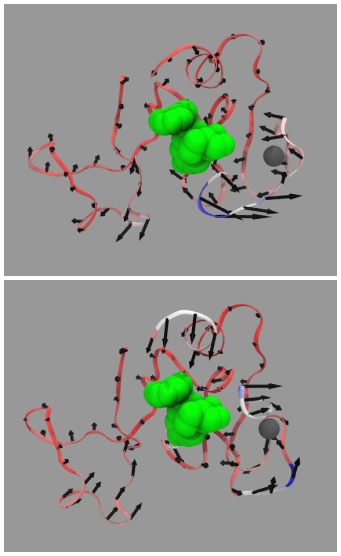

Figure 42: The dominant two PCA modes of the β-pocket. The residues are colored according to their contribution to the PCA mode. The peptide ligand is shown in green in the position in which it occurs in the crystal structure 1FZC. The calcium ion is also shown in its crystal position.

The identified motions show a twist in the CA-loop. Its upper part shows an anti-correlated motion with the loop sitting on top of the coiled-coil region. The glycan binding helix moves perpendicular to that. These motions could play a role in restricting the access of the peptide ligand to the pocket.

The opening and closing was confirmed by a cluster analysis restricted to the C_α-atoms with a DRMS cutoff of 1.35Å. This analysis reveals states in which the pocket shows varying degrees of opening. In the closed states, binding residues can form hydrogen bonds among each other. This essentially only happens between the small central loop (β397–398) and the other three binding residues. These three residues do not form hydrogen bonds among themselves. In the closed state of the unglycosilated simulations, hydrogen bonds between the lower part of the CA-loop and the CC-loop can form between residues β432–433 and β391–392. While these open and closed conformation might restrict access for the peptide, the position of the carbohydrate cluster is much more important for the access to the pocket.

Influence of Glycosilation

The interactions between the glycans and the β-pocket do not leave the flexibility of the pocket unchanged. In Fig. 43, the change in RMSF upon glycosilation is shown. The presence of the glycans is mostly felt in the lower regions. The loop around β397–398 is stabilized by the glycans. Overall the CA-loop becomes stiffer while the CC-loop becomes more flexible. The glycan binding helix is only slightly affected by the presence of the glycans and becomes a little more flexible.

To identify the important configurations of the carbohydrate cluster, a clustering of the pocket is performed including the position of the C_1-atoms of the sugar. Including the sugar increases the DRMS drastically and a cutoff of 3.9Å was used. The pocket shows open and closed states similar to the ones already known. The position of the glycans in these clusters identifies structures (in order of cluster size) where the glycans are associated with the CA-loop, protrude into the solvent or associate

with the coiled-coil region. Only in the case where the glycans protrude into the solvent would the pocket be accessible for the ligand.

The association of the sugars with the CA-loop is stabilized by hydrogen bonds that are present in 43.5% of the simulation time. In contrast to that, hydrogen bonds between the glycans and the CC-loop are only formed in 3.1% of the simulation time and then the glycans almost always (91% of the cases) also form hydrogen bonds with the CA-loop. This disrupts of course the association between the CA- and CC-loops observed in the closed conformation of the unglycosilated pocket. Additionally, this blocks the entrance to the pocket.

In the glycosilated runs, no hydrogen bonds occur between these loops. In the case of the CC-loop, the hydrogen bonds with the glycans merely add to those present in the unglycosilated conformation. For the CA-loop the interactions with the sugar seem to stabilize the loop as more hydrogen bonds within the loop itself are formed in the glycosilated case than in the unglycosilated case.

The sugar groups can also penetrate the binding pocket and form hydrogen bonds with the binding residues. This occurs over 11.3% of the simulation time and involves hydrogen bonds with the binding residues β397, β398, β406. These additional hydrogen bonds mean that the binding residues are more often involved in hydrogen bonding in the glycosilated state. This is not an additive effect however and the frequency with which one binding residue forms a hydrogen bond with another binding residue is reduced. Having the weak association of the sugar with the inside of the binding pocket is not wholly unexpected. In the introduction, it was already discussed that the globular fibrinogen domains evolved from a predecessor that bound sugar groups.

To block access to the β-pocket, the carbohydrate groups do not need to attach to the loops of the pocket itself. Instead, the negatively charged sialic acid groups that terminate the biantennary cluster can interact with positively charged patches in the coiled-coil region. Two such patches are present in the two α-helices formed by the α-chain at this point. Indeed, these are the only patches with which the glycans interact. The patches are formed by the positive residues α138, α141 and α149 on one helix and α183 on the other. Both α138 and α149 are arginine residue while α141 and α183 are lysine residues. The second patch only forms a weak interaction point for the glycans because only one positive residue is present and this is in the immediate vicinity of a negatively charged glutamic acid. A snapshot of this configuration is shown in Fig. 44. In the crystal structure, the carbohydrate groups form a contact with β424 in the CC-loop. No significant interaction between the glycans and this residue is observed in the simulations.

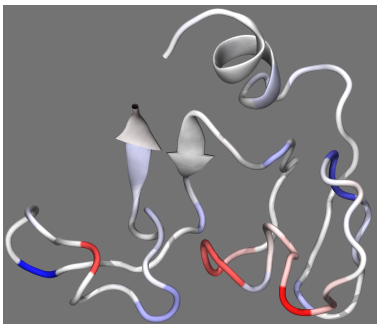

Figure 43: Change of the RMSF of the residues in the β-pocket upon glycosilation. The values range from -1.94Å (red) to 1.98Å (blue).

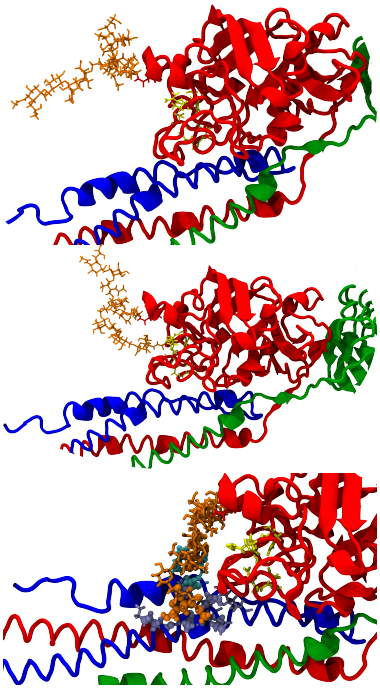

Figure 44: Snapshots of the glycans attaching to the α-chain in the coiled-coil region, penetrating the pocket or protruding into the solvent. Interacting residues of the α-chain are shown with all atoms. Atoms in the carbohydrate cluster and the α chain that are closer than 0.4nm are shown as spheres.

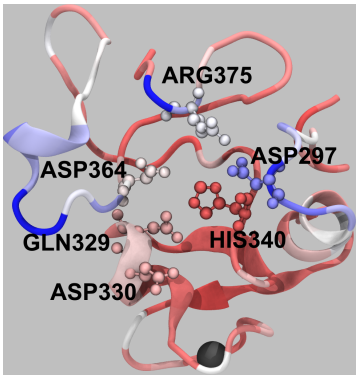

Figure 45: RMSF of the residues in the γ-pocket calculated over all trajectories. The values range from 0.58Å (red) to 5.69Å (blue). Residues with all heavy atoms shown are responsible for ligand binding with the exception of γ375. The calcium ion is shown in its crystal position.

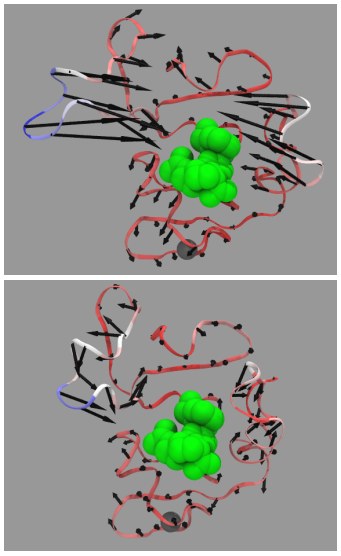

Figure 46: The dominant two PCA modes of the γ-pocket. The residues are colored according to their contribution to the PCA mode. The peptide ligand is shown in green in the position in which it occurs in the crystal structure 1FZB. The calcium ion is also shown in its crystal position.

Known Mutations in Human Fibrinogen

There are no known mutation associated with either the ligand binding residues or the glycosilation site in the β-pocket. Some mutations are known for other residues in the pocket but these exclusively transform glycine to residues with larger side chains so that they likely disrupt the folding of the pocket. Some of these mutated version are not even secreted from the cell in which they are produced [317]

Several mutations are known to occur in part of the FGA gene coding for the charged regions to which the sugar attaches. But only one of these mutations leads to a fibrinogen dimer that is correctly formed. In this variant, the arginine at position α141 is changed to a serine residue. This gives rise to a new glycosilation site and a full bianntenary carbohydrate cluster is attached to the α-chain [306].

5.3.2 *Dynamics of the γ-pocket*

The β-pocket is only important in the lateral association of protofibrils, in the association of the protofibrils themselves the γ-pocket is of greater relevance. To investigate the dynamics of the γ-pocket, the same methods as for the β-pocket can be used and the ligand binding residues have been identified in a crystal structure (PDBID 1FZB) [24, 29]. Because the βC- and γC-domains are homologous, it is expected that some similarities in the dynamics will occur.

Flexible Regions and Modes of Motion: Gating Behavior

The RMSF of the γ-pocket are depicted in Fig. 45. As in the β-pocket, three loop regions show the most movement. None of these regions contain any of the residues for ligand binding. Overall the γ-pocket is more flexible than the β-pocket.

The motions of these flexible parts is characterized by a PCA. The dominant two PCA modes are shown in Fig. 46. The PCA modes are very similar in the glycosilated and unglycosilated case as a PCA overlap of 0.72 indicates.

The PCA reveals that two of the mobile loops (γ374–376, and γ294–299, LIII-loop) show a correlated motion while the third mobile loop (γ354–364, LII-loop) moves in an anticorrelated way. The calcium binding loop (γ318–324, LI-loop) forms a third distinct unit that moves as a rigid region. The large movement of the LII-loop has been noted before and it was termed a movable flap. Likewise, the other regions correspond to the high RMSD regions identified in other simulations of the fibrinogen D-region [252]. The small loop γ374–376 is known to be involved in the D-D-association of two fibrin monomers during the lateral aggregation of protofibrils [62] as well as in the interaction with other proteins (see Table 2).

The principle components of the γ-pocket can be described by an opening/closing of the disordered flaps. This is best characterized by the C_α distance between the residues γ296 and γ358. This distance shows a high correlation with the first PCA modes (correlation coefficient of 0.76 and 0.94 in the case of glycosilated and unglycosilated fibrinogen respectively).

Surprisingly, the gating behavior is not identical in glycosilated and unglycosilated simulations. The histogram of opening distances in Fig. 47 shows a closed state of the pocket in the unglycosilated simulations that is not present in the glycosilated simulations. On the contrary, the pocket seems to have an additional open state in this case. Such an influence of the glycosilation state is surprising considering that the glycosilation site is nearly 7nm away from the γ-pocket.

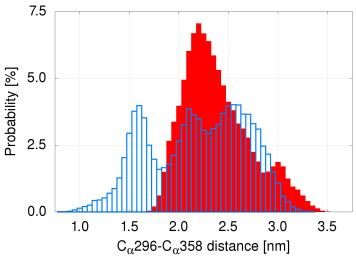

Figure 47: Histogram of gating distance in the γ-pocket for glycosilated (red) and unglycosilated (blue) fibrinogen.

To identify the time scale of the gating motion, a Fourier analysis was performed. As can be seen in Figure 48, the resulting power spectra show an overall power law behavior as is expected from the scale free dynamics of a protein. However, a low frequency peak is also observed indicating a strong periodic motion on longer time scales. The frequency of the anomaly identified by this corresponds to gating motions with a frequency of (4 ± 2)ns in the glycosilated case and (4.8 ± 0.5)ns in the unglycosilated case. Therefore, the gating frequency does not seem to change upon glycosilation.

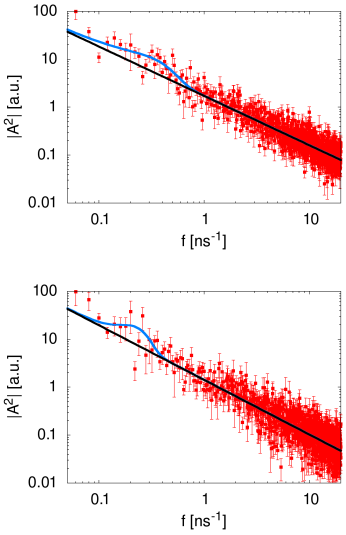

Figure 48: Averaged power spectrum of the C_α296-C_α358 distance fluctuations for the glycosilated (top) and unglycosilated (bottom) simulations

In the LII-loop, a short helical fragment can fold and unfold at residue γ356–358. Because the fragment is only 3 residues long, it can not be classified as an α-helix according to DSSP. Instead, if all three residues are identified as turns the helix is assumed to be formed. This occurs over 37.7% of the glycosilated trajectory and 20.2% of the unglycosilated trajectory. The formation of the helix is not correlated to the opening distance of the pocket (correlation coefficients < 0.1). Examples can be seen in Fig. 49. A similar transient helix is known to occur in the actin D-loop where it changes the interactions between different actin molecules [318]. In a similar fashion the presence or absence of the helix might affect ligand binding or the D-D-interactions in the protofibril.

Interactions of Ligand Binding Residues

To identify important interactions between residues, a cluster analysis of the γ-pocket was performed with a DRMS cutoff of 1.55Å. The cluster analysis reveals multiple open states of the pocket that differ in their conformation of the LII-loop. Closed states are also identified that differ in the position of the side chain of γ375. Some semi-closed states exist where γ375 associates either with the LII- or LIII-loop. Examples of these different states are shown in Fig. 49.

In the glycosilated state, the most populated cluster represents a conformation in which the side chain of γ375 penetrates the pocket and forms hydrogen bonds with the ligand binding resi-

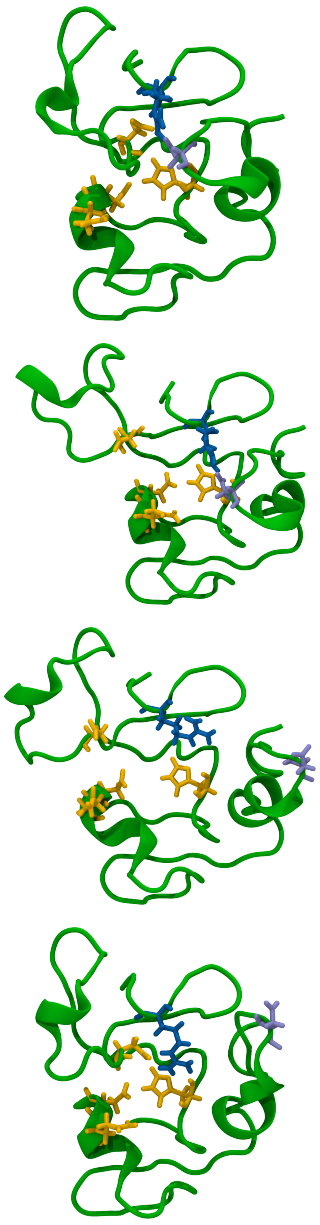

Figure 49: Different conformations of the γ-pocket. The binding residues are shown in yellow, γ375 in dark blue and γ297 in light blue.

dues γ340 and γ364. Residue γ340 has an additional hydrogen bond to γ330 in this case and the pocket is effectively blocked. The second highly populated cluster describes an association of γ375 to the LIII-loop, especially γ297, while the hydrogen bond between γ340 and γ330 remains. The LII-loop is not tightly bound to the pocket in these cases and can open the pocket up. Further clusters describe wide open conformations of the pocket where no hydrogen bonding between the LII- or LIII-loops and the center of the pocket is present. Neither are hydrogen bonds between binding residues.

The relative importance of the different clusters shifts in the unglycosilated case. The most populated cluster is now the association of γ375 and γ297 while the second is an association of γ375 and γ364 that misses the association of γ375 with γ340. Due to the size of the LII-loop, this blocks access to the pocket. A new cluster is identified that corresponds to the closed state of the pocket that is absent in the glycosilated simulations. In this conformation, γ375 forms hydrogen bonds with γ297 and γ364. Thus, both the LII- and LIII-loop are bound to the center of the pocket. A hydrogen bond between γ330 and γ340 is also present so that the pocket is fully inaccessible in this state. Such a closed pocket occurs only over 8.8% of the glycosilated simulation while it occurs in 23.4% of the unglycosilated simulation.

The clusters reveal that an important hydrogen bond is formed between γ330 and γ340. This is present almost entirely throughout the glycosilated simulation (94.1%) but only for half (53.9%) of the unglycosilated simulation time. The change in the hydrogen bonding network is also seen in the reduction of time a hydrogen bond between γ340 and γ375 is formed (from 32.1% to 8.5%) and an increase of time one is formed between γ364 and γ375. In the last case, the hydrogen bonding does not only occur more often but is also stronger. The increase is due to the fact that on top of the primary amines terminating the guanidinium group of the side chain of γ375, the secondary amine is also involved in hydrogen bonds with the carboxylate group of γ364. Stabilization of γ375 by γ364 has been speculated to be important to keep the γ-pocket empty for the ligand to bind [319]. The simulations reveal a more complicated behavior where the pocket can be inaccessible for the ligand even if γ375 is outside.

Even though the association between γ375 and γ340 is weaker in the unglycosilated case, the strong increase in the association of γ364 with γ375 compensates it so that overall the time in which γ375 does not form a hydrogen bond with one or more of the ligand binding residues is reduced by a factor of two. So while γ375 penetrates the interior of the pocket less often in the unglycosilated case the pocket is still less accessible for binding.

Known Mutations in Human Fibrinogen

In contrast to the β-pocket, some mutations have been identified that affect the ligand binding residues in the γ-pocket. As was

mentioned earlier, this does not mean that the mutation rate at this point is higher but rather that a correct functioning of the knob-pocket interactions in the γ-pocket is more important for fibrin formation than in the β-pocket.

No mutations of γ340 have been identified yet but one is known for γ329. This mutation changes the residue from a glutamine to an arginine resulting in impaired fibrin polymerization [320]. The mutation changes the partial charge of the side chain and likely affects the binding of the peptide ligand. In the case of γ330, two mutations are documented. The first changes it from an aspartic acid to a tyrosine residue [321]. The bulkier side chain of tyrosine potentially disrupts the folding pattern so that the ligand can not bind anymore and fibrin polymerization is impaired. The second mutation is less dramatic changing the residue to a valine [322, 323]. This changes its partial charge as well as its potential for hydrogen bonding. This will certainly worsen ligand binding but it can also disrupt the hydrogen bonding to γ340 observed in the simulations. For γ364 it is known that a mutation from aspartic acid to histidine occurs thus possibly disrupting the local folding pattern [324]. As for γ330, a second mutation to valine is also known [325]. Both lead to dysfibrinogenemia with impaired fibrin polymerization. In the case of the mutation to valine, the stabilization of γ375 could not occur. Because this allows the LII-loop to move more, it would open up the pocket. On the other hand γ375 could possibly enter the pocket more and block the peptide ligand. It is unclear what the net effect would be.

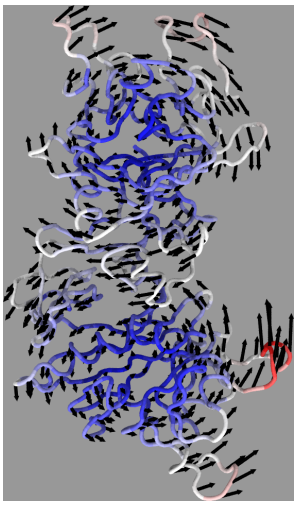

Figure 50: The dominant PCA mode of the fibrinogen D-region is a relative rotation of the βC- and γC-domain. The residues are colored according to their contribution to the mode.

Of the two important non-binding residues γ297 and γ375 identified earlier only γ375 is known to have mutations. A widespread mutation of the highly conserved γ375 residue is its change from arginine to tryptophan [310, 326–331]. This mutation is strong enough to prevent the excretion of fibrinogen from the cell leading to hypofibrinogenemia. Additionally, the degradation of the protein is impaired and the prolonged storage of the mutated fibrinogen leads to liver disease. A second mutation has been identified that changes γ375 to glycine which impairs fibrin polymerization [332]. The results of the simulations suggest that this mutation would remove the hydrogen bonds to the LIII-loop that are likely present in the ligand bound state. The LIII-loop can thus move more freely and is not able to constrain the peptide ligand as well. Surprisingly, it was found that the decreased polymerization of this variant can be counteracted by increased calcium levels.

5.4 INTERMEDIATE SCALE DYNAMICS: DYNAMICS OF THE D-REGION

The dependence of the γ-pocket on the glycosilation state of the β-pocket is surprising considering that they are almost 7nm apart. This long range correlation indicates the complicated cou-

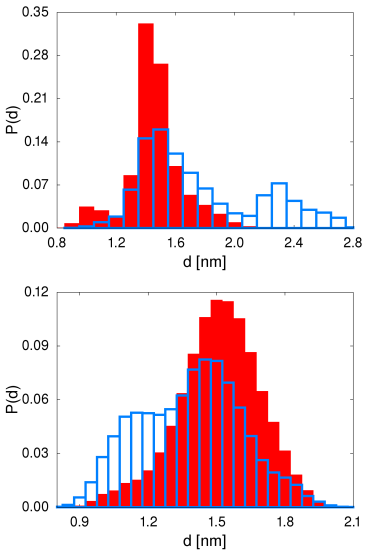

Figure 51: Distribution of the fluctuations of the distance between the P1 binding site and the γ-pocket (top) and β-loop (bottom). The results of glycosilated runs are shown in red, the results of unglycosilated runs in blue.

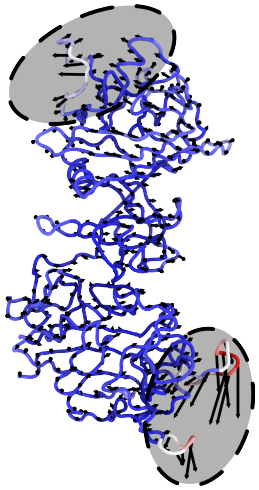

Figure 52: Third PCA mode of the fibrinogen D-region showing a correlated motion of the two ligand binding pockets (highlighted).

pling of different length and time scales in proteins. It is seldom possible to limit the analysis to small subunits and understand the whole picture. Therefore, the dynamics of the full D-region are now investigated. This forms an intermediate level between the large scale hinge motion and the small scale fluctuations in the binding pockets.

5.4.1 *Domain Motion and P1-Accessibility*

The dominant motion of the D-domain is a rotation of the βC- and γC-domain against each other. This is demonstrated by the first PCA mode depicted in Fig. 50. Most of the motion is localized in the ligand binding pockets and the βC-γC-interface. The rotation axes for this motion are perpendicular to the long axis of the D-region and point from the coiled-coil region to the top of the D-region. The second PCA mode is also a rotation of the βC- and γC-domains with rotation axes orthogonal to the short axis of the D-region but perpendicular to the first mode.

The rotations in the first PCA mode result in a periodic narrowing and widening of the cleft formed between those domains. The periodic motion of the cleft opening can best be described by the distance between the C_α-atoms of residues β317 and γ349. This distance shows a correlation of 0.73 with the projection of the PCA mode and allows to estimate the opening and closing time of the cleft using a Fourier analysis. The period is (2.7 ± 0.3)ns for the unglycosilated case and (2.6 ± 0.3)ns for the glycosilated. Only a slight change in the distribution of opening distances of the cleft is observed. At the level of the full D-region, the effect of glycosilation thus begins to vanish.

The P1 integrin binding site is located in the βC-γC-cleft. The opening and closing of the cleft restricts access to the P1 site as is illustrated by a moderate correlation coefficient of 0.6 between the opening distance of the pocket and the number of water molecules within 0.4nm of the P1 residues. The access to the P1 site can also be blocked by the LII-loop of the γ-pocket. However, the loop does not get close enough to interact directly with the P1 residues. Another loop in the βC-domain (β280–285) is also close to P1 and can form hydrogen bonds in rare cases. Hydrogen bonds are present over 10.2% of the unglycosilated simulation time and 6.3% of the unglycosilated simulation time. The hydrogen bonds involve residue β283–285 on the β-loop and γ199 and γ201 of P1. In the unglycosilated case, hydrogen bonds involving γ201 occur much more frequently than hydrogen bonds involving γ199. No such distinction is present in the glycosilated case. In Fig. 51, it is shown that this stems from a tighter association of both loops in the unglycosilated case. Similarly, the closed γ-pocket conformation in this case gives rise to a weaker blocking of P1 by the LII-loop.

5.4.2 *A Long Range Communication Pathway*

The long range correlations evident in the behavior of the γ-pocket could be an indication for a possible allosteric regulation of the β- and γ-pockets. This means binding of a ligand in one pocket would change the affinity of the other pocket for its ligand. In the following, this will investigated.

The first line of evidence for the communication between the binding pockets are correlations between the PCA modes of the pockets. The highest correlations between the PCA modes of the two pockets are obtained for the first mode of the γ-pocket with the the third mode of the β-pocket (0.57 correlation) and for the 8th mode of the γ-pocket with the first mode of the β-pocket (0.60 correlation). No other correlations exceed correlation coefficients of 0.50. The third mode of a PCA of the whole D-region is a mode where most of the motion is located in the movable flaps of both binding sites. The corresponding PCA mode is shown in Fig. 52. No PCA modes showing a correlated motion of the β-pocket with the integrin binding sites can be found. This is however not surprising as the P1 and P2 sites form β-sheets in the rigid γC-domain so that no large scale motion is expected.

This observation highlights correlated motions in the two binding pockets but the question remains how a signal can travel from one pocket to the other.

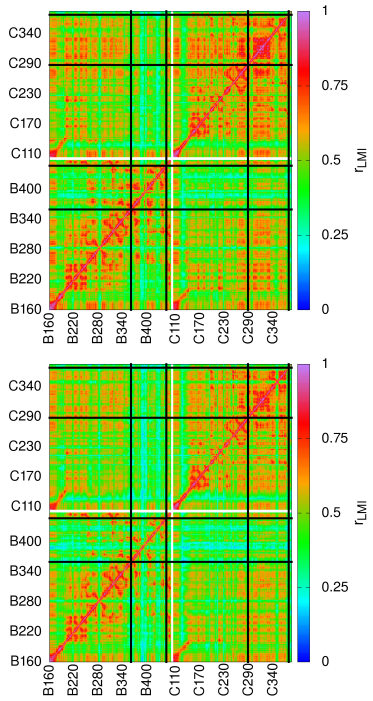

Figure 53: LMI correlation in the fibrinogen D-region for the glycosilated (top) and the unglycosilated (bottom) simulations. The black lines mark the binding pockets.

Identification of a Pathway

To identify highly correlated residues, that could transduce a signal, the LMI correlation matrices for glycosilated and unglycosilated fibrinogen were calculated. Based on this a network was constructed and communication pathways identified as described in section 4.2.2.

The LMI matrices in Fig. 53 show a clear block structure highlighting several secondary structure elements like α-helices and β-sheets. A very prominent feature are the last residues of the coiled-coil region in the lower left of the matrix. These show high levels of intra- as well as inter-chain correlation. Outside of this region no significant correlations of the D-region with the α-chain is found. Furthermore, the LMI matrices reveal highly correlated regions that are not associated with defined secondary structure elements. This is visible in the pockets. As already discussed, both the β- and γ-pocket consist of three larger highly correlated parts most of which are loops. The high correlation between these parts is clearly captured by the LMI.

From Fig. 53 it can be seen that the binding pockets show a direct linear correlation in this picture. A comparison of the two correlation matrices shows that mostly regions corresponding to the two binding sites are affected by the glycosilation. The strong correlation between residues binding the peptides and the calcium binding loops highlights the regulatory importance of calcium ions for fibrinogen function.

The parameter r_{crit}, i.e. the threshold correlation between two residues that are connected, was found to be 0.85 (0.77) for (un-) glycosilated fibrinogen, meaning that a pathway between the two binding pockets exists at a higher level of correlation in the glycosilated case. As a control, pathways between all residues that are at least 6nm apart were identified. The pathway identified between the binding pockets in the glycosilated case does not connect just the pockets. Many more pairs of residues can be connected. Such pathways can in particular be extended to the P1 and P2 integrin binding sites. What is more surprising is that all pathways pass through the same bottleneck of residues at the βC-γC-domain boundary. This bottleneck consists of residues located in the P1-cleft that are contained in all analyzed paths (see Table 14). The bottleneck is bracketed by two funnel regions, that focus existing paths on the bottleneck. The funnel regions are made up of residues containing 25% to 75% of all pathways and residues in their immediate vicinity. The very narrow pathway formed by these residues is shown in Fig. 54. It is interesting to note that residue γ200 is stabilized at the domain interface by a weak interaction between its neighboring residues and a loop in the β region as discussed in section 5.4.1.

Residue	*Fraction*
β204, γ216, γ217, γ225	100%
γ200, γ253, γ348	75%–100%
β211, β225, β249, γ373, γ377	50%–75%
β223, β240, β308, β311, β324, γ341, γ370	25%–50%

Table 14: List of residues that are found in communication pathways in the fibrinogen D-region and the fraction of pathways in which they are contained.

Some emphasis has been placed on the role of hubs [277], i.e. highly connected nodes in the network, and their role for communication pathways. Here hubs are defined as the top 10% of nodes with the most connections (>12 residues with which $r_{LMI} > 0.85$). In the β-chain these hubs are located in the coiled-coil region, as well as in the funnel region for the pathways. In the γ-chain the hubs are located in the coiled-coil region, as well as clustered in and around the γ pocket. The bottleneck residues on the other hand have between two and nine connections and are not considered hubs.

In the case of unglycosilated fibrinogen, r_{crit} has to be lowered to 0.77 to detect a pathway. This allows more diffuse pathways centered around the one identified for glycosilated fibrinogen. However, the residues β204 and γ216 still form a bottleneck and are contained in 97% and 89% of all paths respectively. Some of the bottleneck residues are designated as hubs (now defined as residues with more than 22 connections at $r_{LMI} > 0.77$) in the unglycosilated case. It is noteworthy that residues γ216 and γ217 show almost no increase in connections (see Table 15) indicating a very deep and narrow bottleneck. On the other hand residues β204 and γ225 double or almost triple their connections compared to the glycosilated case. A similar increase in connectivity happens for the residues contained in 75% of the paths in the glycosilated D-region. Similarly, additional hubs are identified that spread the funnel regions in the βC- and γC-domains.

Residue	*w*	*w/o*
β204	5	13
γ216	4	6
γ217	2	2
γ225	9	26
γ200	7	12
γ253	9	29
γ348	5	24

Table 15: Number of connections in the glycosilated (w) and unglycosilated simulation (w/o) of the most important pathway residues.

The identification of the hubs illustrates that the origin of the communication pathway is due to a part in the network that shows only very few connections and is located at the βC-γC-domain border. This bottleneck is surrounded by funnel regions

that collect pathways originating in either domain and focus them on the bottleneck. In combination with the determination of the communication pathways, this highlights the fact that a residue might be critical for signal transduction without being classified as a hub.

A possible mechanism of signal transduction along the pathway would be a disruption of the hydrogen bond network. To check if this is the case, the hydrogen bonding network of the fibrinogen D-region was monitored along the simulation trajectories. The network of persistent hydrogen bonds (hydrogen bonds occurring in more than 25% of all simulation frames) was analyzed for changes upon glycosilation and pathways between the ligand binding pockets.

No long range pathways were found connecting two residues that are at least 6nm apart with persistent hydrogen bonds. Neither does the identified pathway coincide with very stable hydrogen bonds. Similarly, the changes in hydrogen bond network upon glycosilation are not located predominantly on the pathway. However, what can be seen is that many of the changes in the hydrogen bonding network are localized to the binding pockets, highlighting a communication.

The presented simulations do not include the binding of the real ligands, which should be considered in future work. As a substitute, the discussion relied to some extend on the assumption that glycosilation provides enough of a change to trigger an allosteric response. As discussed earlier, the simulations show that the sugars can reach into the β-pocket similar to the fibrinopeptide positioning, but without forming the correct hydrogen bonds. A first step for a computational investigation of the allosteric effect has been done by simulating the pockets with ligands [252]. However, the effect of both ligands being bound simultaneously or in sequence was not considered in this study.

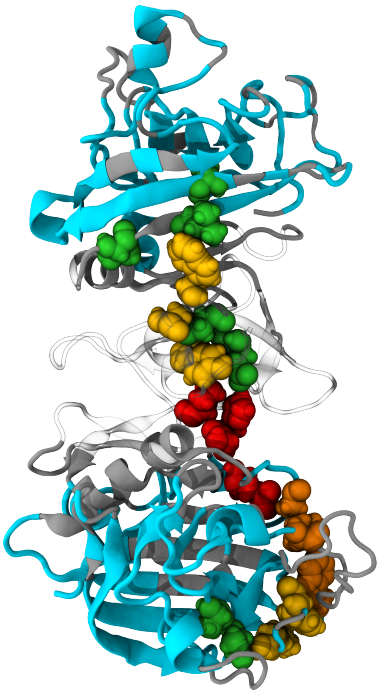

Figure 54: High correlation pathways identified in simulations of glycosilated fibrinogen. Pathways between all residues in the gray regions have been considered. Only residues in the blue regions can be connected. The depicted residues correspond to residues that are in at least 100% (red) to 25% (green) of pathways. See Table 14.

5.4.3 *Evidence for the Existence of an Allosteric Effect*

While the observed direct correlation between the two binding pockets is surprising, the identified high correlation pathway gives a plausible mechanism for signal transduction. Changes in hydrogen bonding seem to be a consequence rather than a mechanism of the pathway. Based on the performed simulations it is not possible to tell the direction in which the signal would have a physiological significance. It is also unclear if its purpose is to regulate the affinity of the ligand or integrin binding sites. Due to the change in opening distances of the γ-pockets, the changes in the hydrogen bond network, and the relatively minor effect on the P1 dynamics it seems more likely that the two pockets are allosterically linked. A possible reason for an allosteric effect is to regulate the correct association of premature fibers. In this hypothesis, binding of the A-knob to the γ-pocket would increase the affinity of the β-pocket for the B-knob. This mech-

anism would support the change in thrombin affinity for FpB cleavage to ensure that lateral aggregation would be much more likely after end-to-end aggregation had already happened. As was found in previous models of fibrin formation, such a delay mechanism is indeed in place under physiological conditions [247].

While there are some investigations into the ligand binding behavior [55, 57, 252, 333], little attention has been paid to the presence of an allosteric effect. In one experiment, the association of the fibrinogen D-region with a fibrinogen fragment containing only the A-knob was investigated with and without a ligand bound to the β-pocket [57]. In this experiment, it was found that the ligand binding in the β-pocket slightly increases the binding of the A-knob fragment. Another experiment looked at the aggregation of fibrin under the addition of different peptides [55]. Unsurprisingly the peptide corresponding to the A-knob fragment inhibited aggregation, while the binding of a peptide corresponding to the B-knob showed an increased aggregation compared to the control. These experimental studies support the long range communication of the two binding sites. In the future, more studies targeted at the change in binding behavior of the peptide ligand could be used to investigate the existence of an allosteric pathway. These studies could use variant fibrinogen with mutations in the γ-pocket that disrupt the hydrogen bonding pattern but not the folded state (e.g [322, 323, 325]) to see if binding in the β-pocket is affected. Alternatively studies of the affinities of deglycosilated fibrinogen should also provide information about the change in binding in the γ-pocket. This would be very similar to the presented simulations and might be more feasible than obtaining variant fibrinogen.

Besides ligand binding studies, other experiments can provide indirect evidence for or against the proposed allosteric behavior. In a deuterium exchange study, it was found that the deuteration dynamics of the γ-pocket changes upon mutation of β235 [37]. This might suggest a different allosteric pathway. In this context, it should be noted that the loop surrounding β235 is a strong connection between the βC- and γC-domains. The hydrogen bond analysis reveals that this loop is the only region that forms interdomainal hydrogen bonds for more than 50% of the simulation time. However, the LMI correlation with the rest of the βC-domain is very weak and no persistent hydrogen bonds are formed. Therefore, even though there is a strong connection to the γC-domain in this loop the decoupling from the rest of the βC-domain prevents the formation of paths. A mutation of this residue, very likely accompanied by structural changes, might open up new pathways by disrupting the hydrogen bonding pattern in wildtype fibrinogen.

Known Mutations in Human Fibrinogen

The mutation of β235 shows that subtle changes likely affect the identified pathway. This is confirmed by a comparison of fibrinogen sequences. With the exception of γ217, all residues contained in 100% of paths are absolutely conserved among the identified fibrinogen sequences. This also holds for residues that are only part of the pathway to a lesser extend (contained in 25%-75% of all pathways). With the three exceptions of residues γ373, β223 and β308 all residues are absolutely conserved, while even those exceptions are conserved within the tetrapods.

The high evolutionary conservation of the residues belonging to the pathway supports their importance for a correct fibrinogen function. Additionally, some mutations leading to fibrinogen related disorders in humans are associated with residues belonging to the pathway or its direct vicinity [296]. In some cases, mutant fibrinogen does not enter the bloodstream so that the effects of the mutations are not characterized and the effect of the mutations remains unclear. In the case of γ200, it has been previously noted that the residue is absolutely conserved, not only in fibrinogen but also other fibrinogen related proteins [334]. The authors speculated that this conservation suggests its structural importance while the current analysis implies an additional functional role. The residues β211 and β240 are cysteines forming disulfide bonds and stabilize the D-region. This is obviously a structural feature that also leads to a high degree of correlation between the residues and explains their evolutionary conservation. All other residues do not have such an obvious structural role.

Mutations of residues close to the identified pathway include β226 that prevents fibrinogen excretion [335] and a mutation of γ227 that leads to an additional glycosilation site [336]. The additional carbohydrate cluster leads to an impaired fibrin formation. A mutation of γ378 is associated with abnormal polymerization and increased break down of fibrin [337, 338]. But the mutation is already located in the γ-pocket so that it likely affects ligand binding more than the allosteric pathway.

A mutation in the γ-pocket that has previously been discussed is the change of γ375 from arginine to glycine [332] . The effects of this mutation can be compensated by increased calcium levels which was puzzling considering the distance between the residue and the γ-pocket calcium binding loop [319]. However, the high correlation pathway would connect both parts and provide an explanation for this behavior.

5.5 SUMMARY

In this chapter, the dynamics of a single fibrinogen protomer and dimer in solution have been investigated. It was found that neither the glycosilation state nor the dimerization state influence the large scale motion of the protomer. This motion has been an-

alyzed using a PCA that identifies a hinge motion. This motion can be characterized by correlated bending and dihedral angles γ and φ. The likely existence of the hinge was investigated for different vertebrate species and the available experimental evidence has been presented. Possible physiological functions for the hinge were proposed.

On a small scale the dynamics of the ligand binding pockets have been studied. The influence of glycosilation on the β-pocket was presented and conformations of the carbohydrate cluster that can hinder access to the pocket have been shown. In the γ-pocket, the importance of the arginine residue $\gamma 375$ has been investigated and different opened and closed states have been identified.

Surprisingly the glycosilation of the β-pocket affects the dynamics of large portions of the D-region including the γ-pocket and the opening of the βC-γC-cleft where the P1 integrin binding site is located. This is explained by a high correlation pathway connecting the β- and γ-pockets. It is possible that this pathway forms the basis for a hitherto unknown allosteric effect between the ligand binding pockets.

6

DYNAMICS AT SURFACES

The study of fibrinogen in solution provided information about the large and small scale dynamics of the protein. But since many experiments have difficulties to access these quantities in solution, they are often performed at surfaces as was discussed in section 2.3. This allows to immobilize fibrinogen and image it more easily. It is therefore of importance to understand the behavior of fibrinogen at surfaces. To some extent, this can be addressed by atomistic simulations. The solution simulations of the last chapter can be used as a benchmark for fibrinogen behavior. This is done here for two model surfaces.

In section 2.3.2, it was discussed that mica and graphite are often used in experimental studies because they allow the preparation of atomically flat surfaces. In section 3.5, it was pointed out that the interactions of both surfaces with biomolecules have also been studied computationally in the past. Because these surfaces are widely used, the initial adsorption stages of fibrinogen at mica and at graphite are investigated here. A short comparison and a discussion about the limitations of these studies follow.

6.1 MICA

6.1.1 *Flexibility and Angle Distribution*

The most prominent motion in solution is the bending at the hinge in the center of the coiled-coil region. This is also evident in the solution simulations. A PCA was performed for each initial orientation separately. The high overlap of the ED subspaces of all three orientations (0.69–0.72) indicates a similar large scale motion. Because the overlap is so high, all orientations were merged and a PCA was performed on the merged trajectories. The dominant modes are bending motions at the hinge and torsion modes in the coiled coil region. This is reminiscent of the behavior in solution.

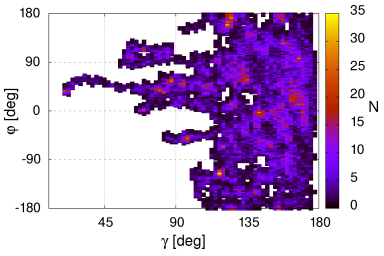

Figure 55: Histogram of the hinge bending and dihedral angles at the mica surface.

A comparison of the PCA modes at the surface and in solution reveals the similarity of the large scale motions. The ED overlap is high (0.63) and increases even (to 0.71) if the calculation is limited to the first 4 modes. Consequently, the large scale motions are similar in solution and at the mica surface. The fibrinogen conformations can be described by the bending and dihedral an-

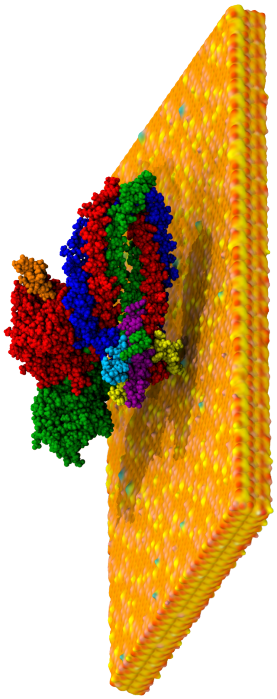

Figure 56: Maximally bend fibrinogen protomer at a mica surface.

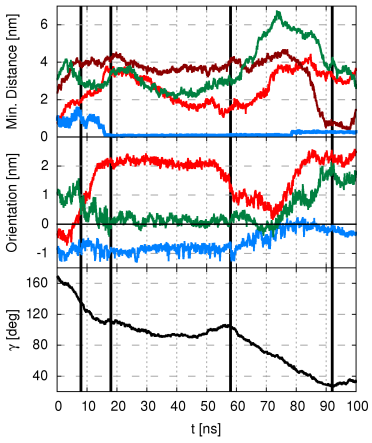

Figure 57: Quantities characterizing the maximally bend adsorption resulting in the configuration in Fig. 56. Quantities for the hinge are shown in green, for the E-region in blue and for the D-region in red (γC-domain in dark red, βC-domain in light red). The black lines separate different stages of the adsorption process.

gles γ and φ. In Fig. 55, the histogram of these angles is shown. The angle distribution is consistent with what has been observed in solution, considering the limited sampling in both cases. The broad distribution of dihedral angles above $\gamma = 120°$ is sampled better at the mica surface than in solution. One of the easy bending directions has not been reached in the surface simulations however.

In contrast to that, the second easy bending direction is sampled to much lower values of γ than in solution. In fact, the sampled conformation corresponds to a maximally bend state. As can be seen in Fig. 56, the angle can not decrease further as the movement is limited by the finite extension of the coiled-coil, as well as the D- and E-regions. It is emphasized that the protein was only confined in the z-direction in this simulation and no collective variable constraints have been applied. The observed strong bending in solution is thus no artifact of these constraints.

The timescale of the bending is similar to what is observed in solution. The average bending time is (16 ± 6)ns with a broad waiting time distribution. The waiting time reaches from less than 1ns to 120ns with a median of 38ns. The bending at the surface happens in competition with the adsorption. Interactions of part of the protein with the surface can limit the bending or lead to rather immobile conformations at the surface. Such cases occur in the simulations and can more than double the bending time.

6.1.2 *Adsorption Dynamics*

That the bending is interrupted by the presence of the mica surface is now illustrated for one case. The graphs characterizing different measures of adsorption and protein orientation described in section 4.2.9 are shown in Fig. 58. A visualization of the different conformations of fibrinogen throughout this adsorption process is shown in Fig. 150 in Appendix B.

The adsorption event starts with an approach of both the D- and E-regions toward the surface. This is driven by the γC-domain that starts out further away from the surface than the βC-domain. After about 6ns, the first contacts are formed between the LII-loop in the γ-pocket and the surface. The movement of the γC-domain past the βC-domain slightly removes the latter from the surface initially but after 8ns the βC-domain also forms contacts. This process is accompanied by a rotation of the D-region with respect to the starting structure accompanied by a slight bending parallel to the surface. At the same time, the E-region approaches the surface without a reorientation. For the rest of the simulation the E-region is immobilized in this position. After 10ns, the D-regions starts to detach from the surface. The γC-domain starts to move away from the surface while the βC-domain briefly forms a tighter attachment through the lysine residue β318 before also moving away. This allows the LII-loop

to make contact again. From this point on the D-region slides on the LII-loop along the surface in a hinge bending motion. During the bending the βC-domain begins to drift away from the surface. At the 30ns mark of the simulation, this motion comes to an end. A further rotation in the D-region sets in. The rotation strains the γ-pocket as the LII-loop tries to stay in contact with the surface. Finally this barrier is overcome and the rotation of the D-region continues, as does the hinge bending. This leads to rolling of the D-region over the surface. But for the limited simulation time, the bending might have continued to conformations with $\gamma < 120°$. The simulation was not continued because a collision of fibrinogen with its periodic image occurs soon after this point. The sliding motion adds another mode of motion of the fibrinogen D-region to the previously identified rolling [232].

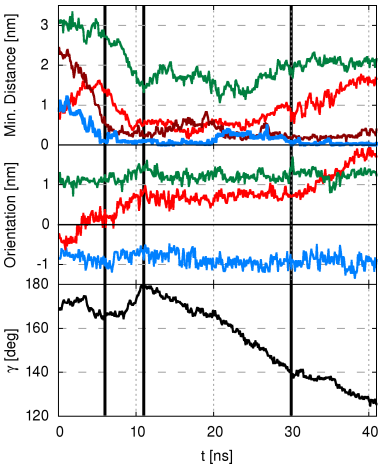

Figure 58: Quantities characterizing an adsorption event of fibrinogen on mica. Color code as in Fig 57.

The example just given illustrates the effect that the somewhat arbitrary choice of $\gamma = 120°$ has for the detection of strongly bent conformations and the calculation of the bending time. As a trend over all simulations, it is found that bending parallel to the surface proceeds unimpeded while bending motions perpendicular to the surface get delayed or even interrupted.

Orientation of the D-region

A similar analysis of the large bending event discussed earlier reveals much more reorientations of the D- and E-region (see Fig. 57). The defined measures can thus pick up orientational changes of the fibrinogen domains during adsorption. The simulations show three distinct states for the orientation of the D-region. Two of these states correspond to orientations in which the coiled-coil region is facing towards the surface or away from it. The third is a side on orientation that does not distinguish if the P1-cleft is facing the surface or away from it. A histogram of the orientation parameter is shown in Fig. 59. Simulations from all three starting orientations show the same three regions of which only two are significantly populated. The less populated region corresponds to the orientation where the top of the D-region points toward the surface. The most populated regions are close to the starting structures in the different initial orientations.

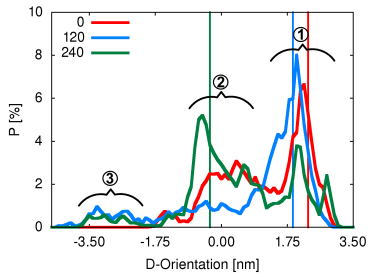

Figure 59: Histogram of orientations of the fibrinogen D-region on mica. The vertical lines indicate the value in the starting configuration.

initial Orientation	*1↔2*	*2↔3*
0	*113/32*	*9/3*
120	*38/13*	*101/25*
240	*13/6*	*78/16*

Table 16: Number of attempted/successful reorientations of the D-region on mica between the states shown in Fig. 59.

Despite the bias toward the initial structures that all distributions carry, the fact that all three orientations produce similar results indicates that the simulation time is long enough to allow for frequent reorientations of the D-region. To quantify such reorientations, transition regions were defined between the three orientations ([1.2nm,1.4nm] and [-2.1nm,-1.9nm]). If a trajectory entered these regions, it was counted as an attempt. If it crossed the region, the reorientation was successful. If it left the region, returning to the starting side for more than 0.15nm, it was counted as a failure. Overall 352 attempts were identified,

95 of which where successful. This corresponds to a successful reorientation event every 22ns with an attempt every 6ns.

The reorientation events are not distributed equally among the different starting orientations as can be seen in Table 16. The 0 starting orientation frequently attempts transitions to a side on conformation but is often not successful. It seems like the 240 orientation is most successful in such transitions. On the other hand the 0 orientation has a higher success rate for transition between side-on and "coiled-coil up" conformations. Both of these effects might well be artifacts of the small number of attempts. In all cases, reorientation events can occur repeatedly between different states.

Most of the final D-region orientations are evenly divided between a "coiled-coil down" conformation (D-orientation >1.5nm) and a side on adsorption. Only two simulations ended in a "coiled-coil up" orientation. Both of these started from orientation 120. In this respect, it is interesting to point out that reorientations to a "coiled-coil up" orientation are attempted more frequently overall but have a significantly lower success rate. Considering the bias of the starting conformations it is likely that the side-on adsorption is a stable adsorption conformation. Which side of the D-region is facing up is not evident from this data and will be discussed later.

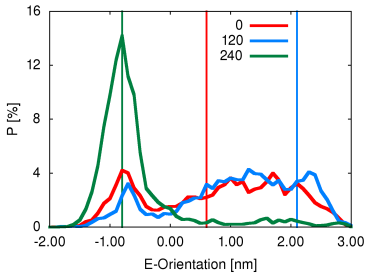

Figure 60: Histogram of orientations of the fibrinogen E-region on mica. The vertical lines indicate the value in the starting configuration.

initial Orientation	*Attempts/ Successes*
0	*81/19*
120	*69/16*
240	*58/16*

Table 17: Number of attempted/ successful transitions between the two conformations of the E-region on the mica surface.

Orientation of the E-region

The two discussed cases of fibrinogen adsorption illustrate that the reorientation of the D-region does not necessarily lead to a reorientation of the E-region. The reason for this is the flexibility provided by dihedral motions of the hinge. This is borne out by the lack of correlation between the orientation of the D- and E-region (correlation coefficient below 0.14). The orientation of the E-region is therefore investigated independently.

The histogram of E-orientations is shown in Fig. 60. Two separate states are revealed. One in which the flexible region of the α- and β-chains point away from the surface ("Fp-up") and one where they point toward it ("Fp-down"). The "Fp-up" orientation is less well defined than the "Fp-down" orientation. The reason for this is that the flexible parts used in the calculation of the orientation can move more in the unrestricted solvent environment to which they are exposed in "Fp-up". In "Fp-down", the Fp-tethers are restricted by attractive interactions with, and sterical hindrance by, the surface.

The initial orientations 0 and 120 are in an "Fp-up" conformation but explore significantly the "Fp-down" conformation as they spend 15% and 25% of their simulation time respectively in this state. The 240 initial orientation on the other hand starts in the "Fp-down" conformation and explores the "Fp-up"conformation in only 9% of its simulation time. This suggest that the "Fp-down" state is the preferred adsorption orientation for the fibrinogen E-region on mica.

Transitions between the two states have been characterized in the same way as for the D-region. Overall an attempt at reorientation is made once every 10ns with a success every 41ns. Reorientations of the E-region are therefore less common than reorientations of the D-region. This finding is surprising considering that the E-region is much smaller than the D-region. The success rate is also much lower than for the D-region. This suggests that the E-region attaches more tightly to the surface and once a suitable orientation is found it is less likely to change.

The transition region was chosen as [-0.25nm,0.25nm]. As can be seen in Table 17, the 240 initial orientation makes the least attempts at reorientation. Because it is almost the same distance from the transition region as the 0 initial orientation, this supports the asymmetry resulting from a preffered orientation. The 120 initial orientation is less likely than the 0 initial orientation to make an attempt at reorientation. This is likely due to its greater distance from the transition region.

	0	*120*	*240*
βC	2%	16%	17%
γC	21%	3%	3%
H	5%	3%	3%
E	35%	31%	47%

Table 18: Fraction of the total trajectory that the different regions of fibrinogen spend in an adsorbed state on mica.

Desorption

The finding that the D-region is more prone to reorientation than the E-region fits the two described adsorption cases at the beginning of the section. In both cases, the E-region attaches early on in the simulations and does not leave the surface again. To see if this is the typical behavior, the number of desorption events was quantified. An adsorbed state was detected if the minimal distance between the protein and the surface was below 0.2nm while a fully desorbed state was identified if the minimal distance was larger than 0.5nm.

The fractions of the trajectory spent in an adsorbed state are given in Table 18. These values highlight that the hinge only rarely adsorbs at the surface. The E-region forms by far the most frequent contacts with the surface. As discussed in the previous section, the initial orientation 240 has the most favorable interaction with the surface.

	0	*120*	*240*
βC	*6/4*	*14/11*	*11/8*
γC	*22/16*	*11/9*	*15/12*
Hinge	*11/8*	*14/14*	*6/5*
E	*13/7*	*19/16*	*13/18*

Table 19: Number of adsorption/desorption events of different regions of fibrinogen events on mica.

The interaction of the globular domains strongly depends on the initial orientation. In the 240 and 120 initial orientation, the βC-domain interacts most frequently with the surface. These frequent interactions are contributed to a substantial part by the sugar groups. For the 0 initial orientation the γC-domain contributes more frequently as in this case the γ-pocket forms contacts.

To highlight the dynamic nature of the adsorption, the number of desorption events are given in Table 19. These numbers show that fibrinogen frequently desorbs from the surface. Less than half of the adsorption events are permanent on the time scale of the simulations. It is striking that the E-region has the lowest desorption rate in the 0 and 120 initial orientations but one of the highest in the 120 initial orientation. This demonstrates the unfavorable nature of the "Fp-up" conformation. The numbers also demonstrate the weak interaction of the hinge with the

Initial Orientation	τ *[ns]*
0	*13±6*
120	*16±4*
240	*10±4*
all	*14±3*

Table 20: Average duration τ between an adsorption and a desorption event.

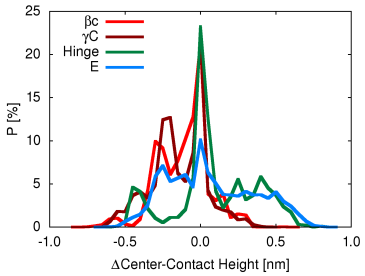

Figure 61: Histogram of the changes in center–minimum contact height characterizing the denaturation of the fibrinogen regions on mica.

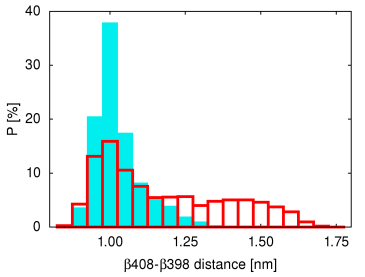

Figure 62: Histogram of the distance between β408 and β398 on mica (red) and in solution (blue).

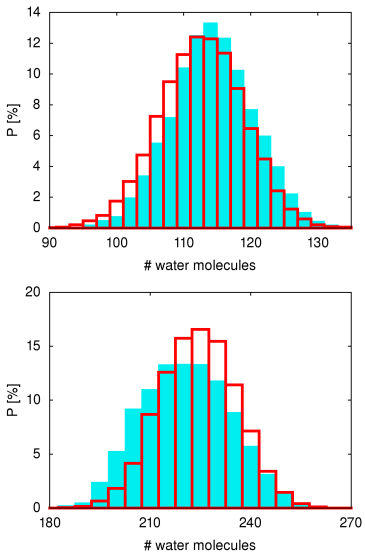

Figure 63: Differences in the accessibility of the P1- (top) and P2-site (bottom) in solution (blue) and at a mica surface (red).

surface. The interactions of the two globular domains with the surface are similar. The βC-domain desorbs less frequently. This is again an effect of the sugar groups that can bridge a substantial gap between the βC-domain and the surface.

After the desorption events are identified, it is possible to calculate the average time spent in an adsorbed state. For this calculation, adsorption events with a duration of less than 1ns were excluded. The results are given in Table 20 and show that the average time a region of fibrinogen spends in an adsorbed state is similar in all initial orientations. Because of the frequent reorientations mentioned early, this is expected. With an average adsorbtion event lasting (14 ± 3)ns, the timescale of the desorption events is thus on the same timescale as the hinge bending. It is also much longer than measured in previous simulations of the adsorption of the γC-domain on polymer surfaces [232].

6.1.3 *Denaturation & Deformation of the Binding Pockets*

The frequent desorption events, even after a significant time in an adsorbed state, suggest that the different regions stay relatively in tact during the adsorption process. This was expected on a hydrophilic surface on the short time scales accessible in the simulations. A histogram of the denaturation distances is shown in Fig. 61.

The results demonstrate that the globular domains show a slight unfolding behavior but mostly remain in their original conformation. The E-region shows larger deformations as the flexible FP-tethers can rearrange more easily. Interestingly the E-region does not only show a decrease but also an increase in height in some cases. This fits reorientations from the "Fp-up" conformation to an "Fp-down" conformation that would lift the center of the E-region. Overall the shape of the protein is mostly conserved in the simulations.

The exposure of interaction sites is an important side effect of the structural changes happening in proteins during adsorption. If the changes affect the ligand binding pockets, they can influence fibrin polymerization. Such changes are revealed by a cluster analysis of the two pockets. As a cutoff, the same value as in the solution simulations was chosen (1.34Å for the β-pocket and 1.55Å for the γ-pocket). The analysis shows only few changes in the γ-pocket. The typical conformations in the γ-pocket are essentially states already identified in the solution simulations. A detectable, but small, difference is a slightly more frequent association of γ375 with γ297.

In the β-pocket, more changes are observed. The least dramatic deformation is a bending of the glycan binding helix in cases where the sugars strongly interact with the surface. Dramatic changes are deformations that pull the CA-loop away from the pocket or the loop around β397 out of the pocket. These deformations can be characterized by the distance of the C_α-atom

of β408 to the backbone nitrogen of β398 and the C_α-atom of β385 respectively. The residue β408 was chosen as a reference because it is located in the center of the pocket that deforms the least. Only the deformation of the loop around β397 occurs frequently throughout the simulation. As can be seen in Fig. 62, a new open state of the pocket is created in the simulations at the mica surface that is not present in solution. It is also observed, that the distance between the CA-loop and the CC-loop is also observed to fluctuate slightly more than in solution.

To determine if the deformations happening upon adsorption can expose binding sites, the water exposure of the relevant residues was measured. This was done for the P1 and P2 sites as well as the binding residues in the β- and γ- pockets. As a measure of exposure, the number of water molecules within 0.6nm of the residues was chosen. According to this measure, the exposure of the P1 site does not change upon adsorption. The exposure of the P2 site increases only slightly from an average of 222 water molecules in solution to 226 in the surface simulations. This is shown in Fig. 63.

A clearer signature of the adsorption is found in the ligand binding pockets (see Figure 64). The exposure of the binding residues in the γ-pocket shows a double peak structure in solution. In the adsorption simulations, only a single peak is visible. This likely reflects a loss of the gating behavior of the pocket. The average exposure of the binding residues is nevertheless similar in both cases. On the contrary, the pulling out of the loop surrounding β397 leads to a higher exposure of binding residues in the β-pocket. The solution distribution of the β-pocket also shows a slight bimodal distribution. The secondary peak corresponding to open conformations is much weaker in solution but strengthened at the surface.

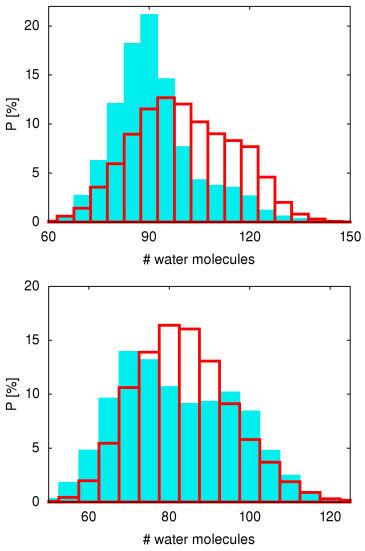

Figure 64: Differences in the accessibility of the β- (top) and γ-pocket (bottom) in solution (blue) and at a mica surface (red).

6.1.4 *Adsorbing Residues*

For the accessibility of the P1-site it is not only important to what degree it is exposed to the solvent but also if the cleft between βC- and γC-domain is facing the surface. This can be clarified by looking at the residues that form the most frequent contacts with the surface.

The total number of contacts formed is given in Fig. 65 as a function of simulation time. It is also shown how the contacts are distributed among the different regions of fibrinogen. Only few contacts are formed at each time and the process of contact formation seems to level of after 30ns to 40ns. It is striking that the E-region contributes much more than would be expected based on the surface area it contributes to the protein. This means the adsorption of the E-region is likely not random. The coiled-coil region contributes about as much as would be expected from random adsorption, while the D-region contributes a little less.

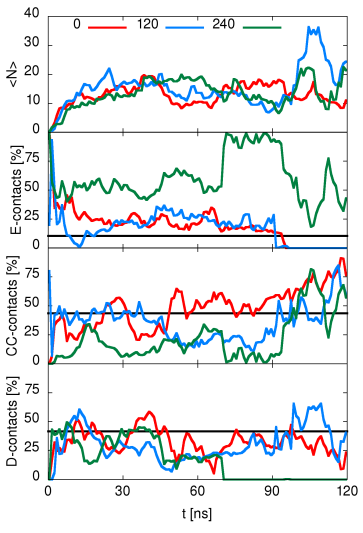

Figure 65: Average total number of contacts formed between fibrinogen and the mica surface as well as the fraction for each region. The black lines correspond to the expected value based on the surface area.

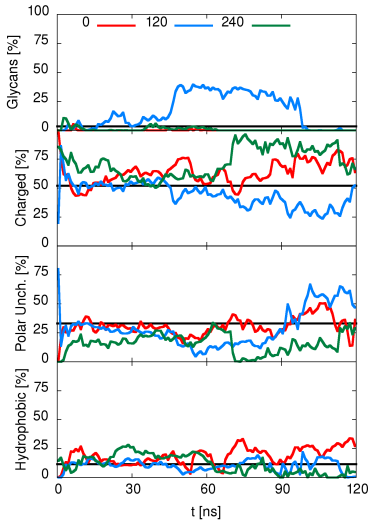

Figure 66: Fraction of contacts formed between different classes of residues and the mica surface.

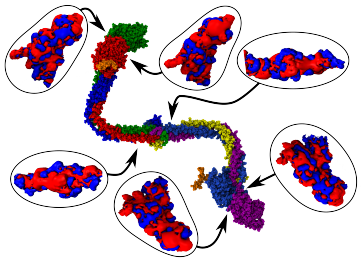

Figure 67: Charged patches in the fibrinogen D- and E-region. The electrostatic potential is shown as isosurfaces at ±26.7mV.

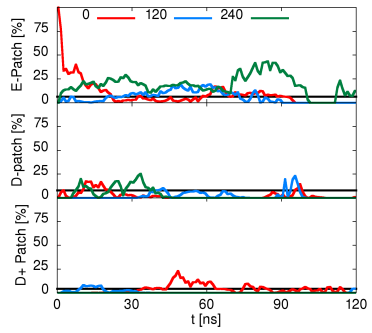

Figure 68: Average fraction of contacts that the charged patches form with the mica surface.

Because the mica surface is charged, a natural guess would be that charged residues interact most strongly with the surface. This is born out by the fraction of contacts they contribute (see Fig. 66). As was discussed in section 2.3, the importance of charged interactions for fibrinogen adsorption is well known [107, 115–118]. The sugar groups do not contribute significantly despite their charged nature. The reason for this is that the terminal sialic acid groups are negatively charged. They also do not contain any carboxylate groups that could trap a potassium ion. Such trappings happen, e.g. for aspartic acid residues. An exception is the 120 initial orientation in which the sugars form many contacts. This can be explained by the orientation of the protein. In the initial orientation, the positively charged patch on the α-chain is pointing toward the surface and one sialic acid attaches there. The second branch of the carbohydrate cluster is then long enough to extend toward the surface and reorient to match the charge structure better. Frequently the sialic acid points away from the surface in these cases.

The presence of extended charged patches would increase the interaction with the surface and could thus potentially form preferred adsorption sites. One such patch was just mentioned and is located in the coiled-coil region where it forms an attachment site for the glycans. Due to sterical restrictions by the stiff coiled-coil region and the rigid βC-domain, it is unlikely that this region will form contacts with the surface. However further patches can be identified in the D- and E-regions as is shown in Fig. 67. The patches were identified from isosurface of the electrostatic potential at ±26.7mV. This is a reasonable value considering that typical values for the ζ-potential of proteins and inorganic surfaces are on the order of 10mV to 50mV under physiological conditions [114, 257]. In this range, the picture of charged patches shown in Fig. 67 does not change much.

The calculation of the electrostatic potential identifies a large negative patch in the D-region. This patch corresponds to the γ-pocket. On the opposite side a small positive patch (D+ in the following) is found that does not correspond to a site with a specific function. In the E-regions, two large negative sites are identified on opposing sides of the coiled-coil region. This is interspersed with small positive regions that are not considered patches.

Overall the charged patches contribute more than expected. However only the E-patch does contribute more to the adsorbing residues than would be expected by chance. This is true at least for the initial adsorption phase. At later times, its significance decreases. The γ-pocket and the D+ patch contribute slightly less than would be expected by random adsorption. The weaker contribution of the patches in the D-region is consistent with the fact that the D-region overall contributes about the expected amounts of contacts. It should be noted that the D+ patch and the γ-pocket can not adsorb at the same time because they are

located on opposite sides of the D-region. From Fig. 68 it can thus be seen that a slight preference exists to have the D+ patch facing away from the surface. Therefore, the charged residues drive the adsorption but the charged patches they form are not overly important for the adsorption.

The contact forming residues can be investigated in more detail to identify a preferred adsorption orientation. The residues forming contacts that lasted at least 1ns where identified in the simulations of each starting orientation. Some of those residues form contacts in all three cases (persistent contacts). This can only be achieved by a reorientation of the corresponding region thus indicating a favorable interaction of this part of the protein with the surface. The residues identified in this way are shown in Fig. 69.

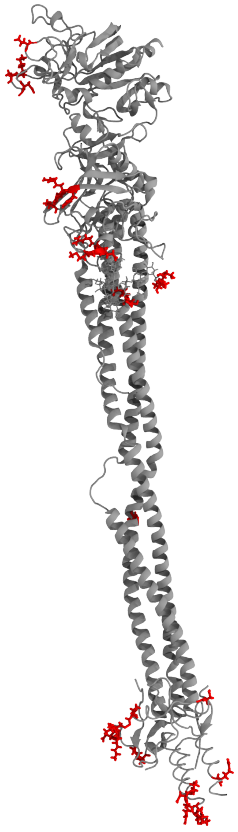

Figure 69: The residues forming persistent contacts with mica (red) independently of initial orientation are located mainly in the E-region and on one side of the D-region.

Two of the persistent contacts are formed by the terminal sialic acids in the carbohydrate cluster. Only one persistent contact is found in the hinge region, while all others are located in the D- and E-region. This highlights the importance of the globular regions for adsorption. It is striking that all persistent contacts in the D-region are formed on one side of the region. This strongly suggests that the adsorption with the P1-cleft facing the surface is preferred as the contacts are formed on this side only. Similarly many persistent contacts in the E-region are located in the flexible Fp-tethers supporting the "Fp-down" orientation. The preferred orientation of the D-region is consistent with the $\alpha_{II}\beta_3$ integrin binding site (located opposite of the P1-cleft) being exposed in the adsorbed state on mica [130].

The persistent contacts formed at the end of the incomplete coiled-coil region are very likely artifacts from terminating the protein structure at this point. This allows side chains from residues on opposite sides of the coiled-coil region to contact the surface. In such cases, the cross-section of the coiled-coil region forms such a small angle with the surface that a full coiled-coil region would intersect the surface. Clearly this is unrealistic.

The residues forming the persistent contacts are almost exclusively grouped around a pair of two oppositely charged residues. For example, the contacts in the D-region feature arginine - glutamic acid or lysine - aspartic acid combinations (see Fig. 70). The interactions of lysine with mica are particularly favorable as the positive charge of the amine is attracted to the negative charge of silicate rings while it can also form hydrogen bonds with the oxygen atoms. In the simulations, it is observed that a single lysine contact can be enough to anchor the E-region. The strong interaction of lysine with silica surfaces has been demonstrated in previous simulations [240]. The positive and negative residues are not necessarily close in sequence as secondary and tertiary structure can bring the side chains close together in the 3D-structure. The feature that allows good adsorption on mica thus seems to be a combination of positively and negatively charged side chains that are grouped close to-

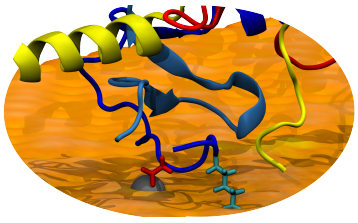

Figure 70: Snapshot of two contacts formed between the fibrinogen E-region and the mica surface. The lysine (light blue) contacts a silicate ring. The aspartic acid (red) traps a sodium ion that has replaced potassium from the counter ion layer.

gether so they can strike a balance between interactions with the negatively charged mica surface and its positive counter ion layer.

Such a combination of charged residues form salt bridges in solution and stabilize the protein conformation. For this reason they are not more frequent on one side of the D-region than on the other. It is thus the question why one side is preferred. A possible reason could be the positive charge of the D+ patch. In the initial approach to the surface, the patch would interact with the counter ion layer and thus experience a repulsion. If the γ-pocket approaches the surface, it experiences an attraction and once it is close enough the positive residues can form interactions with the negatively charged silicate layer. The proposed importance of the negatively charged patches is contrary to what has been proposed in the literature. Due to the negative charge of mica, it was supposed that positively charged patches drive the adsorption [113].

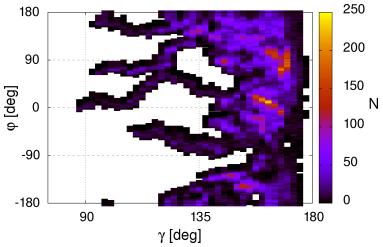

Figure 71: Histogram of the hinge bending and dihedral angles at the graphite surface.

6.2 GRAPHITE

The charged mica surface is an example of a very hydrophillic surface. Depending on the application a more hydrophobic surface might be of interest. To understand the effects at such surfaces, the dynamics of fibrinogen near graphite are investigated. Because graphite has no charged or hydrogen bonding groups, it is also a surface that should show no specific interactions with the protein. It can therefore serve as a control for the observations on mica.

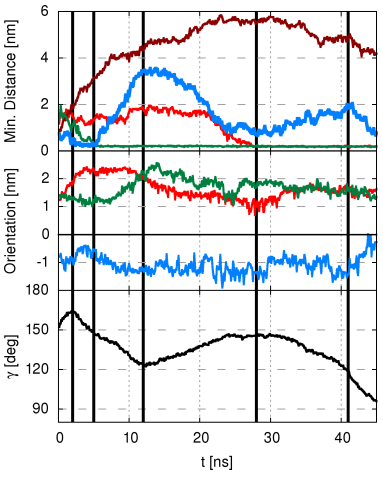

Figure 72: Quantities characterizing adsorption of fibrinogen on graphite. Quantities for the hinge are shown in green, for the E-region in blue and for the D-region in red (γC-domain in dark red). The black lines seperate different stages of the adsorption process.

6.2.1 *Flexibility and Angle Distribution*

The flexibility of fibrinogen is again characterized by a PCA. Like on mica, all three orientations show a similar large scale behavior. This is confirmed by the ED overlap of 0.65. The dominant motions describe a bending at the hinge while higher modes account for torsions of the coiled-coil region. This is similar to what is observed in solution but the ED overlap is slightly worse (0.57) than in the case of mica. This does not increase if the analysis is limited to the first 4 modes. This could be an artifact of the limited sampling time as the longest simulations on graphite are significantly shorter than the longest simulation on either mica or in solution.

The distribution of hinge angles is shown in Fig. 71. The broad distribution of φ is consistent with the other simulations. Also the easy bending directions seem to be consistent. In contrast to the solution or mica simulations, no strong bending events seem to occur at the graphite surface. Considering the short bending time it is unlikely that this is purely a consequence of the short simulation time.

The graphite simulation spend between 3% and 8% of their simulation time in conformations with $\gamma < 120°$. This is considerably less than what is observed on mica. Only four bending events eventually reached conformations with $\gamma < 120°$. The estimated bending time from those events is (17 ± 5)ns which is in line with previously discussed results in solution and on mica.

6.2.2 *Adsorption Dynamics*

It is illustrative to take a look at the adsorption events on graphite to find plausible reasons for the lack of large bending events. The bending angle, orientations and distances of the different regions are shown in Fig. 72. A visualization of the different conformations of fibrinogen throughout this adsorption process is shown in Fig. 150 in Appendix B.

In the first 2ns, the γC-domain moves away from the surface leading to a straightening of the protein. After this, a bending event sets in that proceeds in a plane perpendicular to the surface. The E-region slightly reorients before it briefly attaches to the surface. The hinge moves toward the surface and as it reaches it the E-region desorbs again. This is likely due to geometric constraints. The E-region has a larger diameter than the coiled-coil regions. As a consequence, having both of them attached would necessitate a bending of the stiff coiled-coil region connecting the E-region to the hinge. While the E-region moves away from the surface, a more bent conformation is obtained. In the next stage, starting after 12ns of simulations time, the D-region reorients until the glycans form contacts with the graphite surface. This reorientation reduces the bending again. Finally, a bending motion in a plane parallel to the surface sets in. In the last 4ns, the E-region approaches the surface again while reorienting.

The bending perpendicular to the surface has already been found to increase the waiting time distribution on mica. The difference is that the interactions with graphite are stronger so that desorption events are less frequent. The protein thus rarely gets a second chance at bending.

Surface Mobility: Reorientation and Desorption

The stronger interaction between graphite and fibrinogen is reflected in its decreased mobility at the surface. This is expressed in the histogram of orientations of the D-region in Fig. 73. It is clear from this graph that reorientations are less frequent on graphite as the maxima for each orientation occur close to the starting conformation. Additionally, it can be noted that no well defined states exits as on mica. The maxima of the distributions instead depend mainly on the initial orientation. An analysis of reorientations that is comparable between the three initial orientations is thus not possible. The lack of defined states could be due to the fact that graphite is much stickier for fibrinogen

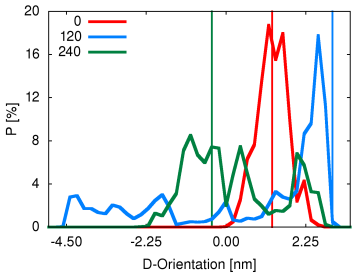

Figure 73: Histogram of the orientations of the fibrinogen D-region on graphite.

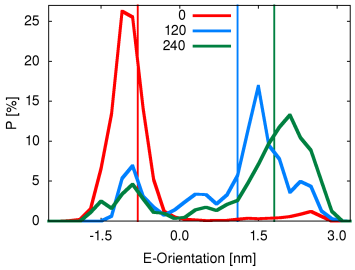

Figure 74: Histogram of the orientations of the fibrinogen E-region on graphite.

initial Orientation	*Attempts/ Successes*
0	*5/1*
120	*12/1*
240	*5/1*

Table 21: Number of attempted/ successful transitions between the two conformations of the E-region on the graphite surface.

than mica. Another, complementary, reason is that no preferred adsorption orientation exists on graphite.

For the E-region the same two states exist that have been detected on mica. The histogram of orientations of the E-region in Fig. 74 shows that all initial configurations explore both states. Both the 120 and 240 orientation on graphite spend 18% of their simulation time in the "Fp-Down" state despite starting in the "Fp-up" state. Only 5% of the trajectory of the 0 initial orientation is spend in this state. Therefore the same asymmetry between the states exists as on mica. A possible explanation for this is that the "Fp-down" conformation is not only energetically but also entropically favored.

However the reorientation events are much less frequent on graphite then on mica. The number of attempts is given in Table 21. Overall an attempt is made every 31ns while a successful reorientation occurs only once in each trajectory. This would translate to a successful event every 229ns but this has little meaning without multiple reorientation events being observed in a continuous trajectory. Likely, a reorientation event happens during the adsorption and the protein is then trapped there.

This is supported by the number of desorption events given in Table 22. Only one of the desorption events occurred after the protein spend more than 1ns in the adsorbed state. This is a much shorter time than on mica and highlights the irreversible nature of the adsorption on graphite.

	0	*120*	*240*
βC	*1/0*	*2/0*	*6/1*
γC	*3/0*	*4/0*	*3/0*
Hinge	*2/0*	*1/0*	*2/0*
E	*5/2*	*3/0*	*3/0*

Table 22: Number of adsorption/desorption events of different regions of fibrinogen on graphite.

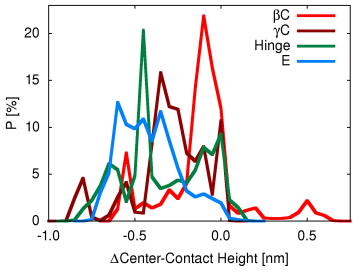

Figure 75: Histogram of the changes in center–minimum contact height characterizing the denaturation of the fibrinogen regions on graphite.

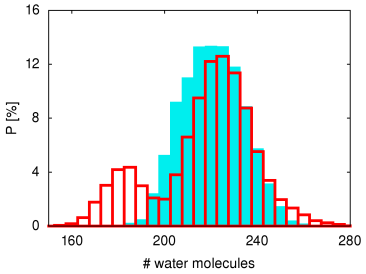

Figure 76: Differences in the accessibility of the P2-site in solution (blue) and at a graphite surface (red).

Denaturation & Deformation

The stronger interaction between graphite and fibrinogen, that makes the surface stickier, should also result in larger deformations. As can be seen in Fig. 75, the change in the diameter of the different regions supports this. All regions show stronger deformations than on mica. Especially, the hinge and the E-region denature much more strongly on graphite because their flexible parts can quickly rearrange. The globular domains are more resistant to these changes and the βC-domain behaves only slightly different on graphite than it did on mica. Most of the deformations correspond to a flattening of the domains at the interface. This can be seen in Fig. 152 in Appendix B.

As a consequence of this behavior, the deformation of the ligand binding pockets is also stronger. The βC-domain shows deformation modes that are similar to those observed on mica. The pulling out of the loop around β397 and deformation of the glycan binding helix are absent. An additional deformation of the CC-loop is found. the main deformation modes in the γ-pocket are deformations of the LII- and LIII-loops. These are pushed away from the calcium binding loop which can, in extreme cases, lead to deformation of the interior of the pocket. The deformations also lead to a slight increase in the association of γ375 with γ297 compared to solution. The association of γ375 with γ340 is slightly decreased.

These deformation impact the accessibility of the binding sites in fibrinogen as can be seen in Fig. 76 and Fig. 77. Both in the β-pocket and the γ-pocket the binding residues become more exposed than in solution. This is caused by the deformation of the loops surrounding the interior of the pocket. The accessibility of the P1-site does not change. the P2-site on the other hand becomes less accessible. This due to the fact that it is facing the graphite surface and all intermediate water molecules are excluded.

6.2.3 *Adsorbing Residues*

If the adsorption on graphite does not lead to a preferred orientation, the initial orientations should not matter for the final adsorption state. Of course this does not take into account geometrical and entropic effects. As can be seen in Fig. 78, only the E-region contributes more than would be expected from a random adsorption. Especially, in the beginning of the adsorption process the E-region is the major contributor of contacts. This is likely due to its flexible parts that can easily adjust to the surface. With time, the contributions decrease in importance. It is expected that over longer times they would reach the value expected from their contribution to the protein surface. The contribution of the E-region is rather indiscriminate of the initial orientation, something that is markedly different from the behavior at mica. From all initial orientations the time the E-region spends in an adsorbed state is around 24%.

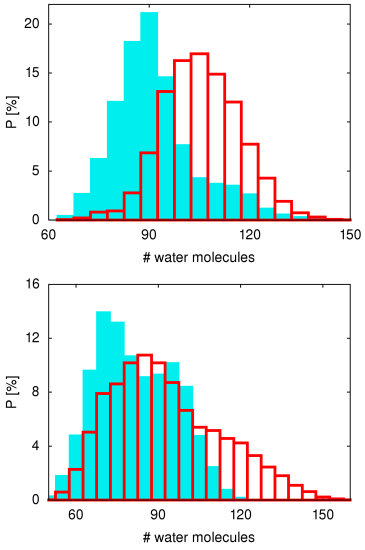

Figure 77: Differences in the accessibility of the β- (top) and γ-pocket (bottom) in solution (blue) and at a graphite surface (red).

The contacts formed by the D-region gain importance as time goes on. The globular domains in this region are more resistant to denaturation. However if they denature they can contribute many more contacts than the E-region because of their size. The γC-domain spends the same fraction of simulation time (30%) in an adsorbed state for simulations started from the 0 and 120 initial orientations. In the 240 initial orientation, it is blocked by the βC-domain that spends the most time (23%) in an adsorbed state in this case. In simulations starting from the 0 orientation, the βC-domain is blocked from the surface by the coiled-coil region and is in an adsorbed state in only 1% of the simulation time.

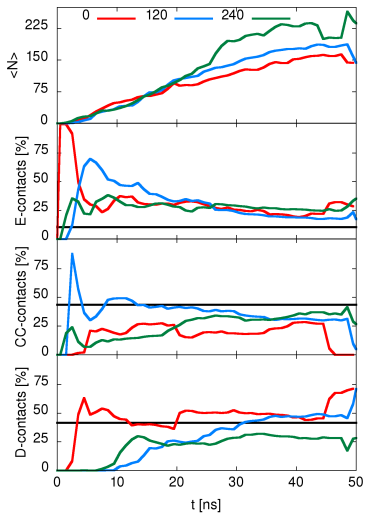

Figure 78: Fraction of contacts formed between different classes of residues and a graphite surface

The coiled-coil region spends the same amount of time (9% of total simulation time) in an adsorbed state independently of initial orientation. It also contributes about as many contacts as would be expected. This is somewhat misleading as the expectation is build on the surface area of the whole coiled-coil region while most of the contacts are formed by the hinge.

The charged patches identified in the previous section can also be considered. The E-patch forms slightly more contacts than expected initially. Over time this relaxes to the expected value. The D+ patch forms contacts only in the 120 initial orientation where it faces the surface. In this case, it contributes the expected

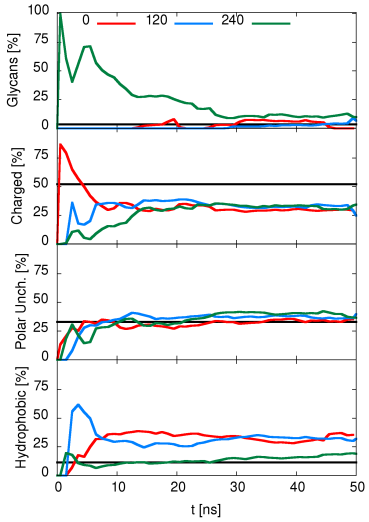

Figure 79: Fraction of contacts formed between different classes of residues and a graphite surface

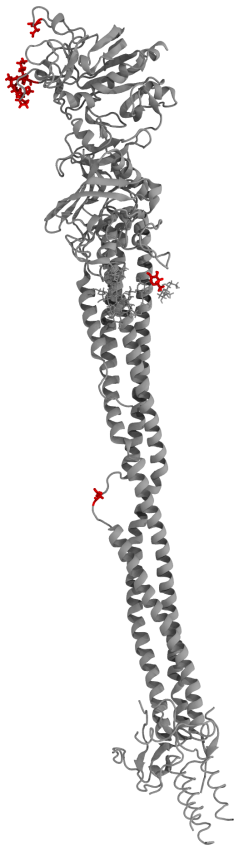

Figure 80: The residues forming persistent contacts with graphite (red) independently of initial orientation.

number of contacts. The γ-pocket contributes significantly more contacts (40% of total contacts) in the 0 initial orientation than expected. As was already mentioned, this orientation is such that only the γC-domain can form contacts in the D-region. In the other orientations, it contributes slightly more than expected with an increase as time goes on. This increase can be explained by the deformations in the loops that bring more residues in contact with the surface.

It is not expected that the charged patches would contribute significantly to adsorption on the uncharged, hydrophobic graphite surface. Their significance for the initial adsorption is instead accidental. The E-region has longer flexible parts that can reach the surface more easily, while the γ-pocket is located in such a way that a bending perpendicular to the interface will bring it in contact with the surface. In Fig. 79, it can be seen that overall the contribution of charged residues is much smaller than would be expected. The significant contributions occur only in the first 10ns of the simulation. After 10ns, an almost steady state for the composition of the adsorbed residues is reached consisting of approximately one third charged, one third hydrophobic and one third polar uncharged residues. This means also that hydrophobic residues are significantly over represented compared to the surface area they contribute. The polar uncharged residues contribute only slightly more than would be expected.

As for mica, the residues forming persistent contacts were identified. The location of these residues are shown in Fig. 80. None of the residues is located in the E-region. A serine contatct is located in the loop responsible for the hinge and another is a galactose unit in the carbohydrate cluster. It is surprising that none of the surrounding sugars or the galactose in the identical second arm of the cluster are identified. That this contact is isolated could thus well be a sampling artifact. All other residues are located in the loops of the γ-pocket. Through bending processes this part of fibrinogen can easily form contacts with the surface.

6.3 COMPARISON AND LIMITATIONS OF ADSORPTION STUDIES

6.3.1 *Comparison of Behavior at Mica and Graphite*

In the last two sections, it has become clear that the behavior of fibrinogen at mica and graphite shows similarities and differences. One of the similarities is the occurrence of the hinge bending close to the surface, as well as the interruption of this process through collisions with the surface. The PCA ED overlap of the mica and graphite systems is 0.65. Thus the hinge bending dynamics at graphite are about as similar to the dynamics at mica as they are to the solution dynamics.

The bending dynamics also immediately demonstrates differences between the two surfaces. While on mica large bending motions are observed, they are absent on graphite. This is attributed to the competition between the irreversible adsorption on graphite and the bending at the hinge. In this respect, it becomes clear that the graphite surface is much more sticky than mica. This is illustrated by the comparative lack of reorientation and desorption events on graphite. Furthermore the number of contacts formed between fibrinogen and graphite is almost an order of magnitude larger than the number of contacts formed between mica and fibrinogen. Overall this means fibrinogen is much more mobile on mica than on graphite. These findings are in good agreement with experiments [110, 112, 127, 129].

A consequence of the stronger interaction between fibrinogen and graphite are the stronger deformations of the ligand binding pockets. While the γ-pocket stays unchanged on mica, it is deformed on graphite. It is interesting that one deformation of the β-pocket is found only on mica. This is the pulling out of the loop around β397 that increases the exposure of ligand binding residues. Because this does not occur on graphite, it seems to be a consequence of specific interactions with the mica surface.

An important result of the simulations at the mica surface was the asymmetry of the adsorbed states with preferred orientations for the D- and E-region. The importance of charged patches was outlined in section 6.1.4. In Fig. 81, the total contribution of the charged patches is summarized for the mica and graphite surfaces. It can be seen that the charged patches contribute more than would be expected from their surface area in both cases. This suggests that these excess contributions do not indicate a specific adsorption mechanism. Rather, they are a consequence of the combination of their position on the protein, hinge bending, and the chosen initial configurations. If a bend fibrinogen molecule would be used for the initial configuration, the contribution of the charged patches should be much less pronounced on graphite. As explained before, their effect for the asymmetric adsorption on mica may be more subtle.

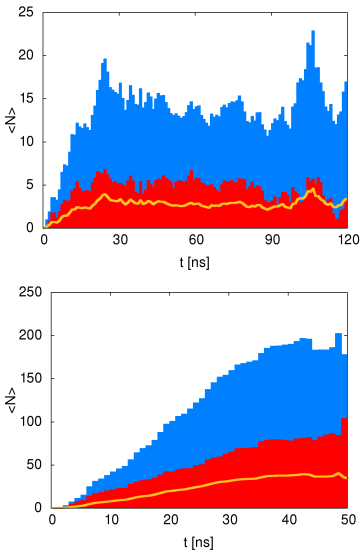

Figure 81: Contribution of adsorbed patches (red) to the average total number of contacts (blue) on mica (top) and graphite (bottom). The contribution almost always exceed the expected amount (yellow line).

6.3.2 *Limitations*

Before the simulation results are compared to experiments, it should be noted that only the very first steps in the adsorption process can be investigated using atomistic simulations. The timescale reached here does not extend beyond the 100ns range while large reorientation events of adsorbed proteins often take minutes or hours [127, 130]. On these longer time scale much more structural changes can happen. Large secondary structure elements can unfold so that, e.g. the helix content of fibrinogen decreases [128]. No structural changes of this extent have been observed in the simulations. This is not surprising considering

that fibrinogen is an extracellular protein and as such built to be resistant to denaturation.

Another limitation are the ideal surfaces used here. While mica and graphite can produce very flat surfaces, they usually contain defects. This was taken into account to some extent on the mica surface where a super cell with a realistic defect pattern was used to build the surface. However this cell is much smaller than the simulation box and an artificial periodicity in the charge pattern is introduced in this way that is likely not present on a real mica surface. Furthermore it is known that the polarizability of materials can lead to effects from deeply buried parts of the surface [237]. The thin slabs of mica and graphite simulated do certainly not capture this.

A more severe drawback of the simulations are the missing parts of the protein. It is known that the αC-regions play an important role in the adsorption and are likely responsible for some of the differences observed for fibrin formation on mica or graphite [27, 144, 145]. Similarly, the preferred orientation of the E-region corresponds to a state in which the Fp-tethers point toward the surface. In the simulations, a large part of these tethers is not included because it has not been resolved in the crystal structure. In reality, this region is therefore much bulkier than in the simulation and might have a different charge structure. it would be surprising if this would not influence the adsorption behavior. The second carbohydrate cluster located in the coiled-coil region is also missing. On mica the adsorption of the hinge seems to be unspecific which could be changed by the presence of an extended, charged structure in this region. It has already been mentioned that the simulation of just part of the second protomer introduces artificial contacts. What their impact on the overall adsorption behavior is, remains unclear.

The limited simulation time is as always an issue. In an equilibrated simulation, the results should not depend on the chosen initial configuration. This is clearly not yet the case for the simulations presented here. The limited time scale also prevents a reasonable estimation of mobility at the surface. Because the largest internal motion (hinge bending) of the protein can not be exhaustively sampled, it makes little sense to try to estimate a diffusion constant from the domain motion as it is still far from random. These restrictions apply even more on graphite where the stronger interaction with the surface slows such processes down even more. Another factor of the limited sampling is that no converged adsorbed state is reached. As such, the calculation of a meaningful adsorption energy is not possible [225]. In principle this would be an interesting quantity for comparison with experiments.

Nevertheless the performed simulations provide valuable insights into the molecular interactions that drive adsorption at the two surfaces. They also provide information about small changes in the ligand binding pockets and a likely asymmetry in the ad-

sorbed conformations on mica that are not accessible to experiments. The implications of such asymmetric adsorbed states on experimental results are explored more in Chapter 8 with the help of a simplified fibrinogen model.

6.4 SUMMARY

The analysis of the adsorption simulations on mica reveal frequent desorption events and a relatively low number of contacts that is formed with the surface. As a consequence, little structural rearrangements are observed. The contacts formed between fibrinogen and mica reveal an electrostatically driven adsorption. Frequent reorientations are observed and highlight the mobility of fibrinogen on the surface. It was demonstrated how a combination of intermediate scale features (charged patches) and small scale features (salt bridges) lead to preferred adsorption orientations of the D- and E-regions. The preferred orientation of the D-region has the P1-cleft at the interface of βC- and γC-domain facing the surface. In the E-region, the Fp-tethers are pointing toward the surface in the preferred orientation. Independent reorientations of the D- and E-region are enabled by the hinge identified in the solution simulations.

The graphite surface is found to be stickier than the mica surface allowing less reorientation and desorption events. At the same time, the stronger interaction leads to the formation of many more contacts. Another consequence of the stronger interaction are deformations of the different regions. These influence the β- and γ-pocket and increase the water exposure of ligand binding residues.

Part III

MODELING

7

A COARSE GRAINED MODEL FOR FIBRINOGEN

The atomistic simulations discussed in the last part can give insights into the small scale dynamics of fibrinogen, be it the hydration of the plasmin binding sites or the behavior of the ligand binding pockets. To a certain extend, it can also illustrate the large scale motions of fibrinogen like its hinge motion. Already at the level of hinge bending, atomistic simulations reach their limit when it comes to accurate determination of the free energy landscape. An equilibrated distribution of γ and φ could not be obtained at this level and more sophisticated sampling techniques like Metadynamics have to be used [339]. Because such enhanced sampling techniques rely on the identification of a reaction coordinate, it is currently not possible to study the aggregation of multiple fibrinogen molecules in atomistic detail. However, enough information has been gathered to construct an initial coarse grained model for fibrinogen based on the atomistic simulations reported earlier. This is done in the spirit of multiscale modeling discussed in section 3.3.1. Because the αC-regions were not simulated, they can not be included in the model using the methodology presented here and will be, for the time being, left out.

In this chapter, a coarse grained model of fibrinogen is developed based on the atomistic simulations presented previously. To avoid artifacts due to insufficient sampling, only the trajectories with more than 40ns simulation time in Table 6 were used. A model is constructed using essential dynamics coarse graining and parametrized using a heterogeneous elastic network. The interaction with other molecules are assumed to be a combination of electrostatics and van der Waals interactions. Parameters for these interactions are determined.

7.1 DEFINITION OF THE COARSE GRAINED BEADS

7.1.1 *Essential Dynamics Coarse Graining*

The first step in defining a coarse grained model is to define the mapping from the fine grained to the coarse grained degrees of freedom. In the present case, this means defining which residues are grouped together into a CG bead. A systematic way to find such a mapping is essential dynamics coarse graining (EDCG)

[193]. The dynamics of a protein can be well described by its essential dynamics, in the following defined as the PCA modes that account for 95% of the correlated fluctuations.

In the CG model, amino acids should be grouped together that move in a correlated fashion. This is achieved by finding a mapping from an atomistic model to a CG model with N_{CG} beads by minimizing

$$\chi^2_{EDCG} = \frac{1}{3N}\sum_{I=1}^{N_{CG}}\left\langle \frac{1}{2}\sum_{i,j\in I}\left|\Delta\mathbf{r}_i^{ED}-\Delta\mathbf{r}_j^{ED}\right|^2\right\rangle \quad (7.1a)$$

$$= \frac{1}{6N}\sum_{I=1}^{N_{CG}}\sum_{i,j\in I}\left[C_{ii}^{ED}-2C_{ij}^{ED}+C_{jj}^{ED}\right], \quad (7.1b)$$

where

$$\Delta\mathbf{r}_i^{ED} = (\mathbf{r}_i - \langle\mathbf{r}_i\rangle)_{ED} \quad (7.2)$$

is the deviation of a particle position from its average projected on the essential dynamics subspace and C_{ij}^{ED} is the covariance matrix of the positions in this subspace. Restricting the analysis to the ED subspace gives more weight to potentially functionally relevant modes and removes high frequency noise.

In the following, only the positions of the C_α atoms are used and it is assumed that each residue is assigned to exactly one CG bead. Minimizing χ^2_{EDCG} will reduce the intra-bead fluctuations and maximize inter-bead fluctuations, i.e. group residues together if their movement is correlated. If MD simulations to obtain the PCA modes are too costly, the analysis can also be performed on an elastic network model of the protein [340]. The problem in this case is that the high frequencies associated with loosely connected regions will lead to a high resolution of these regions despite the fact that they do not have functional importance.

The residual has to be minimized numerically and because the space of mappings is vast a simulated annealing approach is used for global optimization followed by a local steepest descent optimization. In the simulated annealing approach, the mapping is randomly changed and the change is accepted with the metropolis criterion of eq. (3.4) for the residual. The temperature is chosen high at the beginning of the optimization and then successively lowered. In the steepest decent, the boundary of the CG beads are systematically varied by exchanging residues at the interface of two beads. Changes are only accepted if they reduce the residual. Initially, the mappings in EDCG were restricted to mappings that only assigned residues to the same CG bead if they were contiguous in the primary sequence [193]. This approach is known as sequence based EDCG. Later this restriction was relaxed so that residues that are close in space are grouped together (space based EDCG) [341]. The space based EDCG approach with an elastic network model allows it to con-

struct CG models even from low resolution experimental data. It is also useful for complexes with more than one peptide chain.

There is no guarantee that the minimization will find the global minimum and not just a local minimum. To avoid this, many independent minimization runs are started and the results are compared. In this comparison, the consensus mapping of all minimization runs is determined. The consensus mapping is determined by assigning a residue to the CG bead to which it was assigned in a majority of the minimized mappings. This procedure does not necessarily conserve the total number of CG beads. In sequence based EDCG, the comparison of different runs is easy as the sequence of CG beads can not change. Because the bead boundaries can not cross, the initial numbering of the beads is still correct after the minimization. In the space based approach, beads can change their relative position during minimization and it is not clear that bead L in one mapping should correspond to bead L in another. An illustration of this behavior is given in Fig. 82. To nevertheless compare different mappings, it is essential to identify permutations of the bead numbers so that the largest possible overlap in residues assigned to each bead is achieved. For models involving only few CG beads an exhaustive search of possible permutations can be performed but as the computational cost of this rises as $N_{CG}!$ this becomes prohibitively expensive as the number of CG beads increases. A more sophisticated algorithm is described in Appendix C. In this algorithm, all CG beads are renumbered in such a way as to create the maximum overlap regardless of some index occurring multiple times. In a second step, conflicts of multiple indices are resolved systematically. In most use cases, this algorithm should scale approximately as N_{CG}^3.

The only remaining input parameter for the EDCG procedure is the number of CG beads N_{CG}. This number can not be included in the minimization procedure as the residual for proteins has the behavior

$$\chi^2_{EDCG} \propto (N_{CG})^{-2-\Gamma} , \tag{7.3}$$

where Γ is a small, system dependent constant [198]. From eq. (7.3) follows that an increase in the number of CG beads will always reduce the residual. Even worse is the fact, that this happens with a power law dependence so that the behavior is scale free and no optimal resolution exists. It has been suggested to use the number of PCA modes making up the ED subspace as a guideline for the number of CG beads according to

$$N_{ED} = 3N_{CG} - 6 \tag{7.4}$$

but other choices may be justified for other problems.

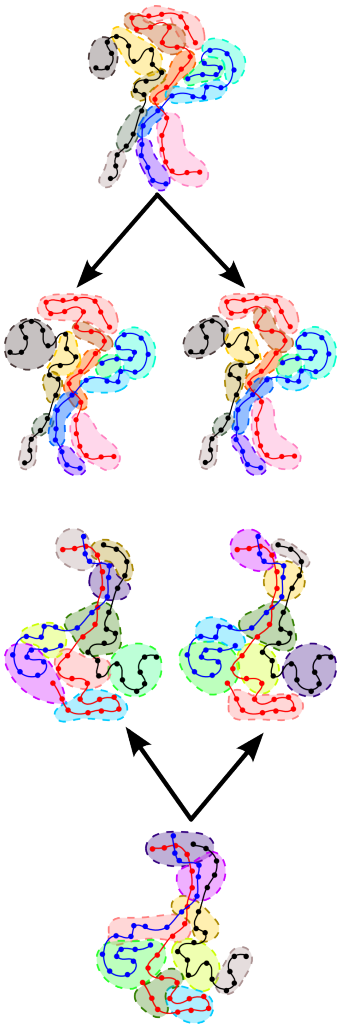

Figure 82: Illustration of two possible results of a sequence (top) and space (bottom) based EDCG procedure. In both cases the two resulting mappings give essentially the same CG model. The models are much easier to compare in the sequence based case as only shifts in the boundary atoms occur while the numbering (color sequence) stays the same.

7.1.2 *Application of EDCG to Fibrinogen*

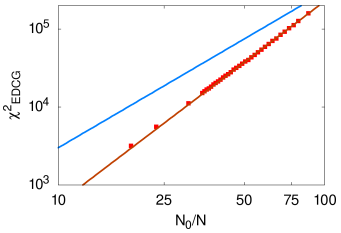

Figure 83: Scaling of the EDCG Residual (equation 7.3) for fibrinogen (red) and N^{-2} behavior (blue) with CG level N_0/N. Here N_0 are the number of C_α atoms in the atomistic representation of fibrinogen .

Figure 84: Assignment of the residues to different CG beads. The residues assigned to a certain bead are shown in the same color.

For the case of fibrinogen the desired model is intended for aggregation studies and it should thus capture the dynamics of the ligand binding pocket in the D-region. In section 5.3 it was shown that each pocket consists of three movable parts. Accordingly, the CG model should represent each pocket with at least three beads. Because this condition is not easily inverted, different models with 12–100 CG beads were studied. This also allows the verification of the scaling behavior in eq. (7.3) that is shown in Fig. 83. The determined value for the correction factor Γ is 0.60 which is in the estimated range of $0 < \Gamma < 1$. A similar scaling behavior has been found for elastic network CG models of fibrinogen [211].

Both space and sequence based EDCG models were considered. For space based EDCG the analysis was at first restricted to the trajectories of the glycosilated fibrinogen dimer, the unglycosilated dimer was excluded because the short trajectories might introduce biases toward straight conformation in the PCA. This analysis could only narrow down the range of total number of CG beads to 39 to 47 CG beads. An exact determination was not possible because the bead assignment in the D-regions was not consistent between minimizations. In the range thus identified the assignment of sites in the two protomers was nearly symmetric and 9 CG beads were located between the hinges in every case.

The symmetry of the fibrinogen dimer, both on an atomistic and a CG level, allows it to consider only the assignment in one protomer so that simulations of the glycosilated protomer and dimer can be combined to improve sampling. On the other hand simulations of the glycosilated and unglycosilated dimer can be combined to improve the sampling in the central region. To implement this, the dimer was split into a central and two outer regions. The central and outer regions where chosen so that they overlapped at the hinge. The central region was defined as the residues from the N-terminus of each of the six chains to (α87,β119,γ62) in both protomers. The outer region was defined as all residues from (α63,β93,γ38) to the C-terminus of each of the three chains. The CG beads for the protomers were thus assigned on data that included the peptide chains starting in the middle between hinge and E-domain, while the beads in the central region were assigned on data that included both protomers up to the hinge. The overlap thus created served as a safeguard to ensure that the assignment in both regions did not conflict. The resulting analysis shows that a model with 20 sites per protomer and 3 sites representing the E-domain consistently assigns 3 CG beads to each of the fibrinopeptide binding sites. The final model was build as a consensus structure of the 10 best results of $1.6 \cdot 10^6$ independent minimization runs. After this, the model was refined by hand to make the mapping continuous along the

structure. It was verified that these modifications reduced the residual. The resulting model has 45 CG beads corresponding to 44 residues per CG bead. The mapping of the amino acid residues to the space based EDCG model is shown in Fig. 84 and its overall structure in Fig. 85.

The space based EDCG model not only reproduces the structure of the fibrinopeptide binding sites but also clearly identifies the hinge as the only part of the coiled-coil region represented by two EDCG sites. One of the beads represents the loop region in the γ-chain responsible for the hinge behavior. The helical segments of the α- and β-chains parallel to it are assigned to the other bead. In the central region the structured elements forming the connection of the protomers by disulfide bonds are assigned two beads, while the flexible parts corresponding to the Fp tethers are grouped into a separate bead. The residues that are assigned to the hinge in this way are α88–99, β120–132 and γ63–75 if only the two beads in the hinge region are taken. If the neighboring two coiled-coil beads are included, the region is expanded to α78–110, β109–142 and γ53–83 which fits the definition based on the atomistic model better. The assignment of the three coarse grained beads to the binding sites corresponds to the division into three parts identified in in the atomistic simulations in section 5.3 as well as indpendent simulations [252] except for some misaligned residues at the boarders of the CG beads. Thus, the model reproduces the binding pockets, the attachment of the knobs, and the hinge and captures the essential features of fibrinogen.

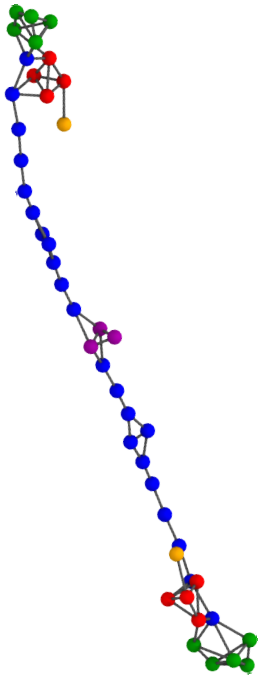

Figure 85: CG model of fibrinogen resulting from space based EDCG. For clarity the different regions of fibrinogen are colored differently and the connection between beads serves as a guide for the eye only.

As a comparison, a sequence based EDCG model has been constructed. The sequence based method can analyze only a single peptide chain at a time so that the trajectories of the individual chains from all simulations could be combined and analyzed. Because the α-chain of fibrinogen does not contain binding sites, the number of CG beads was chosen such that the number of amino acids per CG bead was the same as for the other chains. The result is a model with 58 CG beads. A comparison of the space and sequence based models is shown in Fig. 86. It shows no clear hinge point but the assignment of beads for the binding pockets is similar to the space based model. In the central region, the sequence based model fails to clarify any structural features.

The clear identification of the hinge and the lower overall number of CG beads in the space based EDCG model makes it preferable compared to the sequence based EDCG model for further use. This is not unexpected considering the rigid structures formed by the coiled-coil region that involve multiple peptide chains.

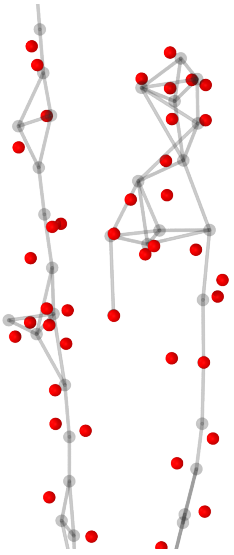

Figure 86: Comparison of the site placement in space based (grey) and sequence based (red) EDCG.

7.2 INTRA-MOLECULAR COARSE GRAINED INTERACTIONS

With a mapping between atomistic and coarse grained model established, the next step is to define interactions between the

CG beads. The large system sizes that are necessary for the simulation of just one fibrinogen molecule prohibit the exhaustive sampling that would be necessary for force matching or iterative Boltzmann inversion. Therefore, the simplifying assumption is made that the interactions can be split up into intra-molecular interactions and inter-molecular, non-bonded interactions. In section 3.3.2, it was discussed that elastic network models have had great success in describing internal motions of proteins. The availability of the atomistic simulations allows the parametrization of an elastic network model for fibrinogen.

7.2.1 *Heterogeneous Elastic Network Models*

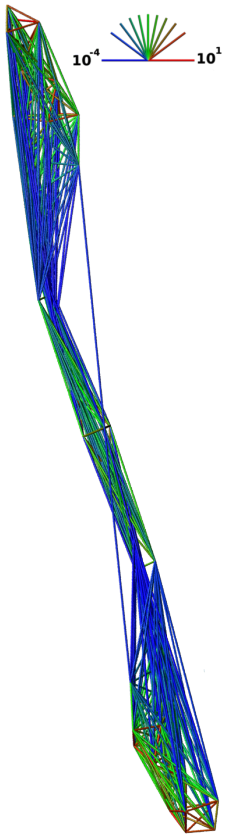

Figure 87: Heterogeneous elastic network for the CG model of fibrinogen. The bonds are colored according to their spring constants in $\frac{\text{kcal}}{\text{molÅ}^2}$ on a log scale.

Traditional elastic network models are usually used to study the dynamics of protein models that are coarse grained on the amino acid level. They connect each C_α atom to others within a cutoff distance using harmonic springs of the same spring constant. In models resulting from EDCG analysis, the different beads might represent regions of different sizes so that the definition of a fixed cutoff is questionable. Furthermore, it has been shown that correlated motions of different parts of a protein depend on a ratio of spring constants and can thus not be captured in an ENM with a single spring constant [342]. These drawbacks can be remedied by using a heterogeneous elastic network model (hENM) [212]. In this approach, initially all CG beads are connected with a uniform spring constant. The equilibrium distance for each spring is set to the average distance between the centers of mass of the amino acids mapped to the bead in a previously performed atomistic simulation. The positions of the CG beads are then relaxed to their equilibrium value using a steepest decent algorithm before the normal modes of the network are calculated. Along the normal modes the fluctuations

$$\Delta x_{ij}^2 = \left\langle \left| \mathbf{r}_i - \mathbf{r}_j \right| - \left\langle \left| \mathbf{r}_i - \mathbf{r}_j \right| \right\rangle \right\rangle^2 \tag{7.5}$$

are calculated. Similar to the EDCG approach this can be restricted to a subspace of the most dominant modes. The spring constants are updated according to

$$\frac{1}{4k_{ij}^{n+1}} = \frac{1}{4k_{ij}^{n}} - \eta \left(\Delta x_{ij,\mathrm{ENM}}^2 - \Delta x_{ij,\mathrm{MD}}^2 \right) \tag{7.6}$$

and the procedure is repeated. The parameter η controls the step size.

This procedure gives a robust way to reproduce intra-molecular fluctuations on the CG level even if the individual bond strength may vary noticeably between different parametrization starting with different initial spring constants [212]. The exact value of the spring constants is thus not essential. What is essential is the topology of the network. The automatic assignment of hard and soft springs, with spring constants varying typically

over five orders of magnitude, is also robust and roughly represents a division into bonded and non-bonded interactions.

Because the fluctuations in an atomistic MD simulation are chosen as a reference point, hENMs are only as good as the simulations used in their parametrization.

7.2.2 *A Heterogeneous Elastic Network Model for Fibrinogen*

The hENM approach is used to construct interactions representing the internal dynamics of fibrinogen. To determine the force constants in the network, first the bonds have to be defined and in the second step, the distance fluctuation between connected CG beads are calculated from the all atom trajectory. Spring constants and equilibrium length are then iteratively adjusted in a third step to reproduce the fluctuations. In all analyses η was chosen as 0.5, only the subspace spanned by the first 135 modes of the elastic network was considered and spring constants were iterated until none of them changed by more than $10^{-4}\,\frac{\text{kcal}}{\text{mol Å}^2}$. This number was also used as a threshold to remove springs if the spring constant was below this value.

To get an unbiased picture of the interactions the analysis is first performed on the trajectories of the glycosilated dimer and all CG beads are connected. This analysis reveals only 509 bonds with a non-zero spring constant instead of the 2025 bonds that would be present if every bead were connected to every other bead. Of these 509 bonds only 10 of the weakest bonds correspond to connections between CG beads above the hinge and CG beads in the other protomer. These bonds are not symmetric between the two protomers and are thus likely artifacts from insufficient sampling. They are subsequently set to zero. This means no bonds between the two outer regions exist. Neglecting the very weak interactions between one of the outer regions and the other protomer allows it to improve the sampling of distance fluctuations within a protomer by including simulations of the glycosilated protomer. Bonds between the protomers only need to be considered for CG beads in the central region. This means that the fluctuations between a pair of CG beads belonging to the same protomer were calculated on the merged trajectories of all simulated trajectories of over 40ns simulation time. To enforce symmetry, these fluctuations were also used for the second protomer. The distance fluctuations for bonds connecting both dimers were checked for symmetry as well. It is necessary to enforce the symmetry at the level of the fluctuations because the interconnected nature of the resulting network makes it impossible to change individual springs later on. Enforcing symmetric distance fluctuations in both protomers thus leads to a symmetric elastic network.

The full elastic network connecting the CG beads is shown in Fig. 87 and consists of 457 bonds with non-zero spring constants. A representation of the model showing springs of differ-

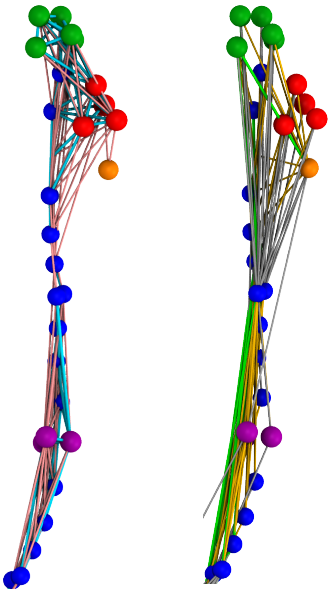

Figure 88: Selective representation of the stiff (left) and soft (right) springs connecting the CG beads in fibrinogen. The bonds are colored according to their spring constant (in units of $\frac{\text{kcal}}{\text{mol Å}^2}$, light blue: $k > 0.1$, pink: $10^{-2} > k > 0.1$, green: $10^{-3} > k > 10^{-2}$, yellow: $10^{-2} > k > 5 \cdot 10^{-4}$, gray: $k < 5 \cdot 10^{-4}$)

ent strength selectively is shown in Fig. 88. This is about half of the bonds that would be present if every site would be connected to every other site. In each outer region, as well as in the central region, it is found that all beads are interconnected. The molecular hinges are therefore the reason for the discrepancy in expected and actual number of bonds. This demonstrates that the hinges act as a bottleneck for bonds and thus separate the dynamics of the different regions. In Fig. 87 it can also be seen that the strong bonds form the backbone of the model, similar to the connections shown in Fig. 86. The weaker bonds on the other hand form usually longer range connections. An interesting small scale feature revealed by the CG model is that the sugar groups have a closer affinity to the calcium binding loop than to the other side of the β-pocket.

7.3 INTER-MOLECULAR COARSE GRAINED INTERACTIONS

The intra-molecular interactions are completely covered by the elastic network model and for multi-protein simulations it is assumed that only CG beads of different fibrinogen molecules interact via non-bonded interactions. Similar to atomistic force fields it is assumed that these non-bonded interactions can be captured by a Lennard-Jones potential responsible for excluded volume interaction and short ranged attraction, and an electrostatic part.

7.3.1 *Lennard-Jones Type Interactions*

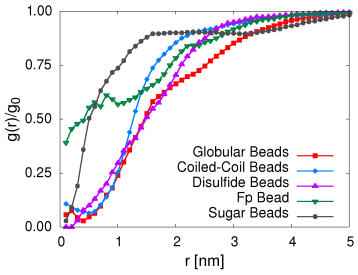

Figure 89: Distribution of water around the different classes of CG beads

To calculate the interaction of different CG beads with each other simulations of fibrinogen aggregates would be necessary as a target for parametrization as discussed in section 3.3.1. Because this is presently not feasible, a different approach is taken to determine the parameters for the Lennard-Jones potential. The first parameter is the bead size and can be estimated based on the excluded water volume.

To determine the excluded volume of the CG beads the distribution of water molecules around the protein in the atomistic simulation was calculated and averaged. The resulting distributions were averaged and grouped into five categories: Coiled-coil beads, beads belonging to the globular domains, beads representing sugars, beads representing the disulfide knots in the E-domain, and the bead representing the flexible termini in the E-domain. These categories are based on the shape of the water distribution as well as the location of the bead in the protein. To obtain a simple model, only one excluded volume interaction will be introduced for each category. To this end, the water distribution functions in each category are averaged over all beads in this category. The resulting distribution functions are shown in Fig. 89. It is clear that no attractive interaction between the CG beads and water is present at this level of resolution. A draw-

back of this simplification is that the hinge region will be more extended than the rest of the coiled-coil region.

To obtain the radius of each bead the potential was calculated using the Boltzmann inversion in eq. (3.8). It was found that the a purely repulsive Morse potential of the form

$$V(r) = D \exp\left[-2a\left(r - r_0\right)\right] \tag{7.7}$$

describes the potentials better than the repulsive part of a Lennard-Jones potential. Because this procedure only aims to estimate the size of the beads, the repulsive Morse potential is used for fitting. The values determined for r_0 are given in Table 23.

With the bead size determined, the well depth of the Lennard-Jones potential remains to be estimated. Due to the number of amino acids contained in each bead a simple interpretation of this parameter in terms of hydrophobic or hydrophilic beads is not possible. Rather this parameter should contain information about specific interactions between different parts of the protein that become important when two or more fibrinogen molecules interact. This can be illustrated for the case of the interaction between the beads representing the γ-pocket and the Fp-tether bead. In this case the well depth should account for specific hydrogen bonds and possibly small scale structural rearrangements that stabilize the Fp-tether ligand in the pocket. Similar stabilizing interaction come into play for the other beads in the D-region to accurately capture longitudinal and lateral association. As the coiled-coil regions do not contribute to specific interactions during fibrin assembly it might be appropriate to model them as purely repulsive beads.

Bead Type	r_0 *[nm]*
Globular	*17.2*
Coiled-Coil	*12.7*
Disulfide	*14.8*
Fp	*7.3*
Sugar	*6.9*

Table 23: Estimated radius of the CG beads determined by fitting a repulsive Morse potential to the water radial distribution function.

In a study of HIV capsid protein, this was done by building CG models of experimentally known structures and reducing the well depth until these structure became unstable [343]. While this potentially over stabilizes experimental structures, it would be a way to estimate these parameters. Unfortunately structures on the resolution of the present CG model are rare. On an atomistic scale only X-ray structures of the D-D interface are available and AFM and electron microscopy data do not have a high enough resolution to parametrize the CG beads. Another approach could be to iterate the parameters until correct protofibril structures are identified. Considering the missing pieces in the atomistic and thus the CG model, as well as the computational effort of such a fitting procedure, this does not seem like a realistic approach. At the moment, the Lennard-Jones well depths thus remain undetermined.

In other highly coarse grained models like shape based Coarse Gaining [344] the Lennard-Jones radii have been determined by the radius of gyration of the underlying structure. The well depth was set to a value of 4kcal/mol in this case. As this does not reproduce the specific interaction of ligands and pockets that is desirable for association studies, such a fixed value can only be a first step.

7.3.2 *Electrostatics*

To parameterize the electrostatic part of the inter-molecular interactions it is necessary to define a systematic way to assign a charge to each CG bead. At present, no dipole or polarizability will be included for the individual beads. Due to the size of the particles, it might not be advisable to simply use the sum of the amino acids assigned to a bead. Such a model would only reproduce the monopole moment of the electrostatic potential of each bead. For a better representation it would be advantageous to instead reproduce the total electric field of the protein as best as possible. Such a method is presented here.

Derivation of the Method

Most protein simulations are performed in an aqueous solvent with physiological salt concentrations. In such an environment, screening effects play an important role. The screening length λ is typically less than 1nm, e.g. in a 1M NaCl solution at a temperature of 310K the Debye screening length is $\lambda_D = 0.7735\text{nm}$. Hence, a screened potential of the form

$$\phi(\mathbf{r}, t) = \sum_i \frac{q_i}{4\pi\varepsilon_0 |\mathbf{r} - \mathbf{r}_i(t)|} e^{-\frac{|\mathbf{r}-\mathbf{r}_i(t)|}{\lambda}} \tag{7.8}$$

is more appropriate at the targeted CG level. It is assumed that the value of the screening length is the same in the atomistic and the CG system. A more detailed derivation for vacuum electrostatics as well as for the case of mixed vacuum electrostatics (for the atomistic model) and screened electrostatics (for the CG model) can be found in Appendix D. In the following, capital letters will refer to CG coordinates and charges and lower-case letters to atomistic quantities.

To produce the best match between atomistic (AA) and CG potential, the residual

$$\chi^2_\Phi(\{\mathbf{r}_i\},\{\mathbf{R}_I\},\{q_i\},\{Q_I\}) = \left\langle \int d^3r (\Phi^{AA}(\mathbf{r}, t) - \Phi^{CG}(\mathbf{r}, t))^2 \right\rangle_{AA} \tag{7.9}$$

is minimized with respect to the CG charges Q_I. The average is performed over the simulation trajectory. In this way the effect of different conformations is taken into acount with the appro-

priate statistical weight for a given temperature. The integration in eq. (7.9) can be performed explicitly and gives

$$\begin{aligned}
\int d^3r\, \Phi^{AA}(\mathbf{r},t)\Phi^{AA}(\mathbf{r},t) &= \sum_{i,j=1}^{N_A} \frac{q_i\, q_j}{(4\pi\varepsilon_0)^2}\, f(\mathbf{r_i}(t),\mathbf{r_j}(t)) \\
\int d^3r\, \Phi^{CG}(\mathbf{r},t)\Phi^{CG}(\mathbf{r},t) &= \sum_{L,M=1}^{N_{CG}} \frac{Q_L\, Q_M}{(4\pi\varepsilon_0)^2}\, f(\mathbf{R_i}(t),\mathbf{R_j}(t)) \\
\int d^3r\, \Phi^{AA}(\mathbf{r},t)\Phi^{CG}(\mathbf{r},t) &= \sum_{i=1}^{N_A}\sum_{I=1}^{N_{CG}} \frac{q_i\, Q_I}{(4\pi\varepsilon_0)^2}\, f(\mathbf{r_i}(t),\mathbf{R_I}(t)),
\end{aligned}$$

with

$$f(\mathbf{x},\mathbf{y}) = \int d^3r\, \frac{e^{-\frac{|\mathbf{r}-\mathbf{x}|}{\lambda}}}{|\mathbf{r}-\mathbf{x}|}\, \frac{e^{-\frac{|\mathbf{r}-\mathbf{y}|}{\lambda}}}{|\mathbf{r}-\mathbf{y}|}. \tag{7.11}$$

In these calculations, the periodic boundary conditions in the simulation were ignored because the screening length is so short that the box might as well be infinite. In general a simulation box should be constructed in such a way in any case to avoid finite size effects.

Choosing $\mathbf{x}$ to be the origin and expressing the integral in spherical coordinates, eq. (7.11) can be written as

$$f(\mathbf{x},\mathbf{y}) = 2\pi \int_0^\infty dr\, r e^{-\frac{r}{\lambda}} \int_{-1}^{1} d\tau\, \frac{e^{-\frac{\sqrt{r^2+a^2-2ar\tau}}{\lambda}}}{\sqrt{r^2+a^2-2ar\tau}}, \tag{7.12}$$

with $a = |\mathbf{x}-\mathbf{y}|$. Using the substitutions $\alpha = r^2 + a^2$, $\beta = 2ar$, and $\kappa = \sqrt{\alpha-\beta\tau}$ the integral can now be easily solved and gives

$$\begin{aligned}
\int_{-1}^{1} d\tau\, \frac{e^{-\frac{\sqrt{\alpha-\beta\tau}}{\lambda}}}{\sqrt{\alpha-\beta\tau}} &= -\frac{2}{\beta}\int_{\sqrt{\alpha+\beta}}^{\sqrt{\alpha-\beta}} d\kappa e^{-\frac{\kappa}{\lambda}} \\
&= \frac{2\lambda}{\beta}\left(e^{-\frac{\sqrt{\alpha-\beta}}{\lambda}} - e^{-\frac{\sqrt{\alpha+\beta}}{\lambda}}\right) \\
&= \frac{\lambda}{ar}\left(e^{-\frac{|a-r|}{\lambda}} - e^{-\frac{a+r}{\lambda}}\right).
\end{aligned} \tag{7.13}$$

With this, one finds

$$\begin{aligned}
f(\mathbf{x},\mathbf{y}) &= \frac{2\pi\lambda}{a}\int_0^\infty dr\, \left(e^{-\frac{|a-r|+r}{\lambda}} - e^{-\frac{a+2r}{\lambda}}\right) \\
&= 2\pi\lambda e^{-\frac{|\mathbf{x}-\mathbf{y}|}{\lambda}}
\end{aligned} \tag{7.14}$$

and can write down an expression for χ^2_Φ. In general, one has to check if the existence of χ^2_Φ implies further conditions. As a consequence of the long range interactions, this is the case for vacuum electrostatics where χ^2_Φ only exists if the total CG charge is equal to the total atomistic charge. Constraints of this kind can be treated with the help of Lagrange multipliers. For the short ranged screened electrostatics no additional constraints are necessary on mathematical grounds, nevertheless they can

be included. However, this might not be advisable on higher CG levels. If the size of the beads is large enough, it is likely that a bead would encompass not just the amino acids assigned to it but also a counter ion layer. In this case, enforcing the charge conservation may worsen the results.

The CG charges are now determined by minimizing the residual, i.e. solving

$$\frac{\delta \chi^2_\Phi}{\delta Q_I} = 0, \forall I, \tag{7.15}$$

which can be written as a matrix equation:

$$\underline{\underline{A}}\,\underline{Q} = \underline{\underline{B}}\,\underline{q} \tag{7.16a}$$

$$A_{ij} = \left\langle -\frac{\lambda}{\pi N_t} \sum_{n=1}^{N_t} \exp\left[-\frac{\left|\mathbf{R}_i(t_n) - \mathbf{R}_j(t_n)\right|}{\lambda} \right] \right\rangle_{AA} \tag{7.16b}$$

$$= A_{ji}$$

$$\underline{\underline{A}} \in \mathbb{R}^{N_{CG} \times N_{CG}}$$

$$B_{ij} = \left\langle -\frac{\lambda}{\pi N_t} \sum_{n=1}^{N_t} \exp\left[-\frac{\left|\mathbf{R}_i(t_n) - \mathbf{r}_j(t_n)\right|}{\lambda} \right] \right\rangle_{AA} \tag{7.16c}$$

$$\underline{\underline{B}} \in \mathbb{R}^{N_{CG} \times N_A}.$$

These are high dimensional equations but can be solved efficiently using numerical tools. The position of the CG beads is defined as the position of the center of mass of the residues mapped to the bead.

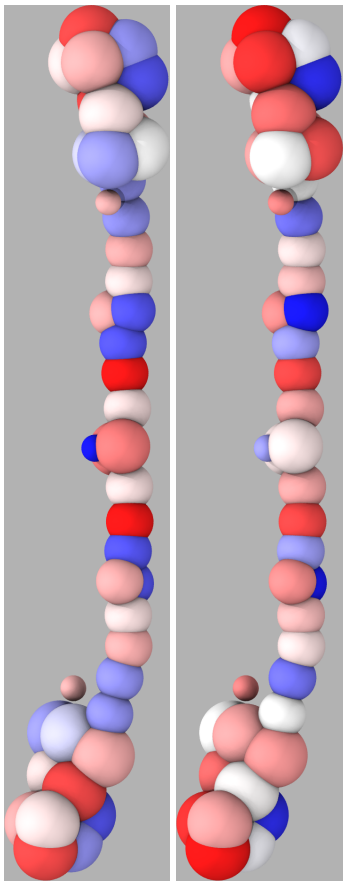

Figure 90: Representation of the non-bonded CG interaction of fibrinogen. The CG beads are drawn according to their excluded volume radius and colored according to their charge (from red representing the most negative to blue the most positive values). Left: Method (i), Right: Method (ii).

Charge structure of the Fibrinogen Model

To test the validity of different approaches in determining the CG charges they were determined with four different methods:

(i) A simple summation of the charges of all the atoms assigned to a CG bead

(ii) Charge fitting using a screened potential without charge conservation

(iii) Charge fitting using a screened potential with charge conservation

(iv) Charge fitting using an unscreened potential with charge conservation

Because vacuum electrostatics is a long ranged interaction and the D-domains carry non-zero net charges, it is not possible to include protomer simulations to improve the sampling. Only the trajectories of the glycosilated dimer were used. The charges resulting from the analysis are already very symmetric despite the limited sampling. Typical variations are less than 15% between protomers. To obtain a symmetric model, the charges of the CG

beads in the protomers are averaged. The resulting charges can be found in Table 29 in Appendix D.

The resulting overall charge structure is similar in all cases and the results are shown in Fig. 90 and Fig. 91. The obtained charge pattern corresponds well to the results of the APBS calculations in section 6.1 with a large negative patch on one side of the D-region and a small positive one on the other. They also consistently show a dipole for the hinge beads. Only the charges resulting from method (i) and (ii) reproduce positively charged Fp beads. This is an important feature to capture the effects of electrostatics in fibrinogen aggregation. In this regard, method (iii) does the worst job in recovering the local charge structure represented by method (i). It has many more negative beads than should occur. Because getting the wrong sign for the charges is not just a matter of scaling the potential but changes which beads are attracted, this would have a major impact on aggregation. In this respect it is interesting to note that method (i) gives a different result than the charge fitting methods (ii)–(iv) for the sign of 3 of the 4 beads representing the βC-domain. The charges in the βC-domain also provide a hint regarding differences in the higher order moments of the field generated by the different approaches. Using method (i), the bead representing the glycans has the exact opposite charge of the bead representing the protein region to which the glycans are attached. Thus, a strong dipole moment will be present in this region. In the potential fitting procedure both beads have negative charges giving rise to a completely different field.

The charge fitting procedure should also be applicable in the development of a CG model for the mica surface. In this case problems arise because the mica slab extends to the border of the simulation box and the periodic boundary conditions have to be respected. Possible ways to do this are outlined in Appendix D.2.

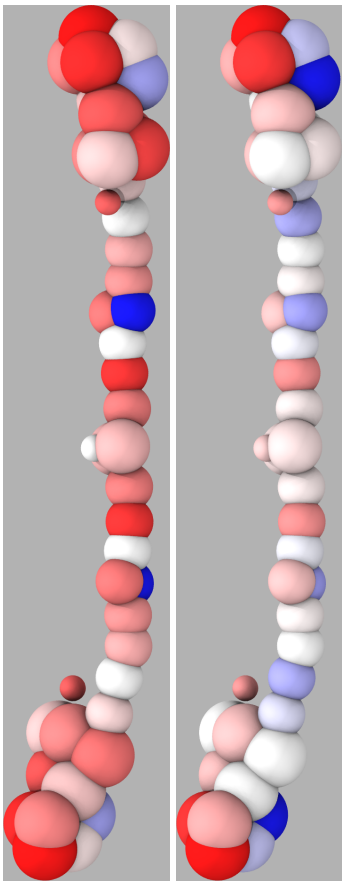

Figure 91: Representation of the non-bonded CG interaction of fibrinogen. The CG beads are drawn according to their excluded volume radius and colored according to their charge (from red representing the most negative to blue the most positive values). Left: Method (iii), Right: Method (iv).

7.4 SUMMARY

In this chapter, a coarse grained model for fibrinogen was developed on the basis of atomistic simulations. The model structure was derived using a space based EDCG method and beads were connected by a heterogeneous elastic network. The model reproduces relevant features of the protein such as the hinge and the binding pockets. The electrostatic interactions were parametrized by a newly developed method to fit the electrostatic potential of a CG and an atomistic model. An attempt was made to parametrize a Lennard-Jones potential to complete the non-bonded interactions of the CG model. An estimate for the bead radii could be found but insufficient data was available to parametrize the interaction strength.

As all of the parameterizations are done with respect to the atomistic simulations, their accuracy is limited by the limited

sampling of the atomistic phase space. Another major hurdle resulting from the limited atomistic simulations is the lack of a representation for the αC-regions or the Fp tethers. Both are important for protofibril formation [345] and should be included in a model that wishes to accurately describe fibrinogen aggregation.

8

A PHENOMENOLOGICAL MODEL FOR FIBRINOGEN

In the first part of this thesis, the large scale dynamics of fibrinogen have been investigated using atomistic detail and in the previous chapter a CG model was developed to simulate aggregates. Both of these approaches retain many details that might not be necessary for a conceptual understanding of fibrinogen adsorption. For this reason a simple, phenomenological model (as opposed to the systematically coarse grained model of the last chapter) is developed that can capture the behavior of fibrinogen at surfaces.

As was discussed in section 3.5, up to now only one model for the adsorption of fibrinogen was available that tried to reproduce the correct molecular shape of fibrinogen. This model was developed by Adamczyk, Barbasz and Cieśla and consists of a string of rigidly connected beads of different diameters [254, 255]. The model thus has a dumbbell shape and was studied with a random sequential adsorption model. In this model, individual fibrinogen molecules are irreversibly placed on the surface in a successive, non overlapping way. An important finding of these studies was that the dumbbell shaped model can cover a larger fraction of the surface than a simple cylinder of comparable dimensions. But this fraction was still significantly less than what is observed in experiments [255]. To remedy this, an end-on adsorption mode was introduced in addition to the side-on adsorption based on experimental data [134]. In this end-on adsorption, fibrinogen extends far into the solution. The observed flexibility at the hinge makes it seem unlikely that fibrinogen would stick in this manner into the solution. One could nevertheless expect that due to this flexibility fibrinogen can cover a larger surface fraction without resorting to end-on adsorption. Thus, the results of the atomistic simulations can improve this model. In the following section an improved model will be presented and compared to experiments.

8.1 THE MODEL

The phenomenological model of fibrinogen is meant to reproduce only the large scale behavior and consists of five points representing the D- and E-regions as well as the hinges. The hinges and the E-region are connected by a single rigid rod, while the

hinges are connected to the D-regions with rods that can pivot around the hinges. In the atomistic simulations, a striking similarity between the large scale motion (PCA overlap and angle distributions) in solution and at surfaces was observed. This similarity makes it unlikely that interactions with a surface distort the hinge motion beyond recognition. The D- and E-regions are modeled as hard spheres and the finite extension of the coiled-coil region is incorporated by spheres of the appropriate diameter at the hinge points. A schematic of the model is given in Fig. 92. To allow the definition of dihedral angles for each protomer a dummy point x_0 is defined close to the E-region. The surface is represented as a hard wall. In this way it only acts as an obstacle and allows to study the effects of fibrinogen flexibility on adsorption without specific interactions.

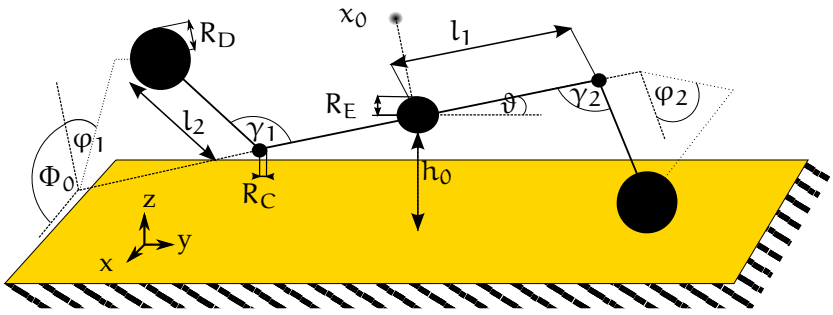

Figure 92: Phenomenological model for fibrinogen at a simple hard wall surface.

	length [nm]
l_1	*8.49*
l_2	*10.61*
R_D	*2.35*
R_E	*2.44*
R_C	*1.02*

Table 24: Size Parameters for the phenomenological model in Fig. 92.

Fibrinogen can approach the surface in different orientations. This includes rotations around its long axis, as they were discussed for the orientational screening in section 4.1, and a tilting with respect to the surface. The first effect is included by introducing an angle Φ_0 between the plane described by $\varphi_1 = 0$ and the surface. The tilting is described by an angle ϑ between the surface and the central rod.

The dimension of the model components can be extracted from the crystal structure and are given in Table 24. The length l_1 and l_2 are measured as the distances between the centers of the E-region (defined as all residues from the N-termini to α58, β90 and γ31 in both protomers), the hinge region (as identified in section 5.2: residues α98–α109, β136–β142, γ68–γ79) and the D-region (defined as all residues in the B and C chain from β200 and γ140 to the C-termini). Although neither the D- nor the E-region is really spherical, they are approximated this way. In this rigid model, the extension of the E-region along the long axis of fibrinogen is not relevant, so that the diameter in a plane orthogonal to this axis is used to determine R_E. The same holds for the D-region. For the later the extension in the directions orthogonal to the main axis of fibrinogen differ by about 1nm and their arithmetic mean is used. A case where the oblate forms of the globular regions becomes important are potential collisions between them. In reality, these are not only determined by the interaction of the residues in the domains themselves, but also

by the interaction of their hydration shells. This leads to complicated interactions, that depend on the exact form of the domains as well as the solvent. The phenomenological model will neglect these effects and approximate the globular domains as hard spheres. The diameter R_C of the spheres at the hinges is determined as an average of the distances from a heavy atom in the center of the coiled-coil region to the heavy atoms in the same plane (orthogonal to the long axis of fibrinogen) that are away furthest from the center.

The probability distributions for the hinge movements can be extracted from the simulation. Alternatively, a model distribution can be used to investigate the sensitivity of the model to certain parameters. Based on the atomistic simulations in section 5.2, it is assumed that the two hinges are independent. Furthermore, the bending and dihedral angle are correlated and are described by a joint probability distribution $P(\gamma, \varphi)$. As a first approximation the distribution extracted from atomistic simulations in solution is used. For a more accurate model the correct free energy landscape of the hinge should be determined.

To compare the model to experimental adsorption studies, it needs to be defined what is meant by an adsorbed state. In the following an adsorbed conformation is defined as one where the minimal distance between the spheres representing the D- and E-regions and the surface is smaller than some distance h_{max} (0.05nm in the following). This amounts to the assumption that the coiled-coil regions do not contribute significantly to the adsorbtion energy of the real system. This is motivated by the fact that the globular regions are able to deform more than the rigid coiled-coils. Therefore, more residues can come in contact with the surface and increase the interaction. This assumption is supported by AFM measurements that indicate that fibrinogen is adsorbing mainly through its D- and E-regions [125].

Experimentally the coiled-coil regions are not resolved, when statistics about the bending are collected. They are therefore allowed to fluctuate in this model via a variation of ϑ and h_0. The presence of the hard wall introduces constraints between the different possible parameters in the model. Additional constraints result from the hard sphere interactions between different parts of the protein model. A derivation and discussion of the constraints can be found in Appendix E.

8.2 COMPARISON TO EXPERIMENTS

In experimental studies only few characteristics of the adsorbed conformations can be collected. These are usually limited to the contour length of the adsorbed molecules and in some cases the D-E-D bending angle α shown in Fig. 93. This angle can also be extracted from the model.

Let $\mathbf{r}_{1/2}$ be the vector from the center of the E-region to the center of the left/right D-region. Then α is given by

$$\alpha = \cos^{-1}\left(\frac{r_{1,x}\, r_{2,x} + r_{1,y}\, r_{2,y}}{\sqrt{(r_{1,x}^2 + r_{1,y}^2)(r_{2,x}^2 + r_{2,y}^2)}}\right). \tag{8.1}$$

The usage of $\cos^{-1}(\ldots)$ implements the experimental bias that, because the connector regions are not visible, no values $\alpha > \pi$ can be measured. Because fibrinogen is flexible in solution and at surfaces, no single value of α is representative. Instead the distribution $P(\alpha)$ needs to be compared to experiments. As is discussed in Appendix E no analytical expression for $P(\alpha)$ can be derived and even simple numerical schemes are not promising. However, the limiting cases discussed in the appendix strongly suggest that in experiments the cis conformation shown in Fig. 93 is observed as well as the trans conformation.

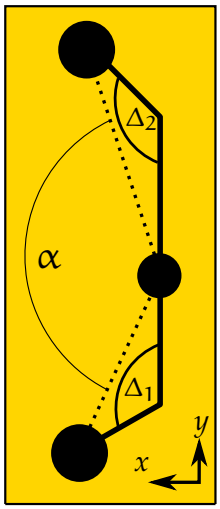

Figure 93: Top view of an adsorbed phenomenological model. All quantities are calculated from the projection of the molecule on the xy-plane.

8.2.1 *Computational Study of the Model*

Despite the difficulties in deriving an expression for $P(\alpha)$ a comparison with experiments is desirable. To this end, a Monte Carlo (MC) approach is taken. Such an approach has the added benefit that it can be extended to a full MC simulation of the phenomenological model that includes more than one fibrinogen molecule as well as adsorption and desorption events. However, to investigate effects like competition between bending and crowding in a simulation, the free energy surface of the hinge needs to be determined.

In the MC approach, random conformations are generated by systematically varying h_0 and Φ_0 (see Fig. 92) and drawing random angles ϑ, γ_i and φ_i. The hinge angles are drawn from the distribution observed in the simulation, and ϑ is sampled from a uniform distribution. The program proceeds as follows:

1. Draw N^* random conformations for each pair (h_0, Φ_0) and count the $N_A^{(h_0,\Phi_0)}$ conformations that satisfy the constraints (see eq. (E.2) in Appendix E).

2. Calculate the total number of accepted conformations

$$N_A = \sum_{h_0,\Phi_0} N_A^{(h_0,\Phi_0)}$$

   and the weighting factors $w(h_0, \Phi_0) = \frac{N_A^{(h_0,\Phi_0)}}{N_A}$.

3. Randomly draw $w(h_0, \Phi_0) \times N$ allowed conformations for each (h_0, Φ_0) and save them.

With this method, on average N fibrinogen conformation are sampled for each (h_0, Φ_0) pair, but the contribution of each such pair is weighted according to the accessible conformation space.

A histogram of the bending angles or the resulting distribution of dihedral angles can be determined from the saved model conformations.

The described weighting procedure is necessary because in the case of flat adsorbates (i.e. $h_{max} \to 0$) the allowed regions in the conformation space become one dimensional lines, due to the hard body interactions. Accordingly, the probability of drawing a valid conformation at random approaches zero. In this case, even large values of N^* will result in only a handful of valid conformations due to the high rejection rate. The typical number of conformations obtained in this way are far to small to obtain accurate statistics for the behavior of the model. The weighting procedure is essentially an importance sampling procedure similar to the Metropolis criterion discussed in section 3.1.1. It focuses the sampling on regions where possible conformations can exist and reduces the sampling in region where this is not possible.

To speed up calculations, the simulated angle distribution was approximated by a series of two-dimensional Gaussians. To compare to experiments it has to be taken into account that the sampling of the hinge free energy surface is far from converged. The relative height of the population maxima for the hinge is thus very uncertain and likely biased by the starting conformation. To account for this, the distribution $P(\gamma, \phi)$ was divided into seven regions representing certain population maxima (see Fig. 94). It was assumed that the position of the population maxima is correct but their height was adjusted to best reproduce experimental results. This was done by representing the probability distribution for an angle α as

$$P(\alpha) = \sum_{ij} a_{ij} P_{ij}(\alpha) \tag{8.2}$$

where $P_{ij}(\alpha)$ is the distribution of α if one hinge is restricted to region i while the second hinge is restricted to region j. The individual $P_{ij}(\alpha)$ are shown in Fig. 157 in Appendix E. The matrix of fitting parameters is necessarily symmetric and can be decomposed as $a_{ij} = a_i a_j$ for independent hinges. An advantage of this procedure is that it allows the introduction of correlations between the two hinges if the decomposition is not done.

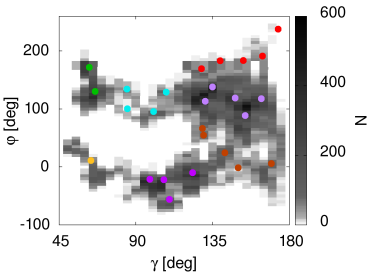

Figure 94: Divison of the hinge distribution into seven regions (colored). The distribution $P(\gamma, \varphi)$ is approximated by 2D Gaussians (centers shown as points).

8.2.2 *Comparison on Mica and Graphite*

At this time only three experiments have been published that measured the distribution of the bending angle α [17, 64, 143]. Of these only one is done on a well defined surface, namely mica [17]. The other experiments were performed on graphite that has been made hydrophillic using glow discharge [143] and graphite that has been covered by a proprietary amphiphilic glycine-carbohydrate modifier [64]. The distributions $P(\alpha)$ on mica and the modifier covered graphite are very similar. Because

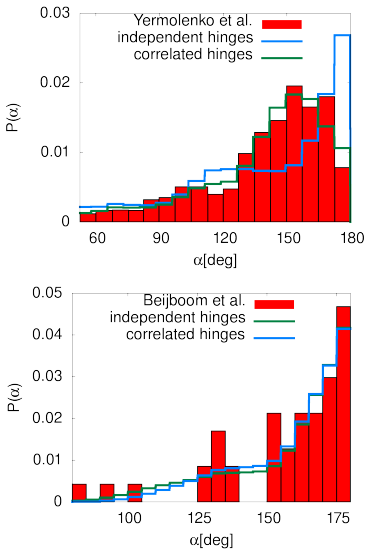

Figure 95: Results of the fit of the phenomenological model to data collected on mica (top) [17] and graphite (bottom) [143]

the structure of the modifier is not disclosed and self assembly processes of the modifier have not been studied it is hard to speculate about the surface structure in this case. However, assuming cyclic forms for the carbohydrates and the possible presence of charged groups it could be that the resulting surface is similar to the mica surface. The distribution on hydrophilic graphite is different but has a worse statistic.

To determine the a_{ij} a least squares fit was performed to the experimental data. In the correlated model the number of parameters would be greater than the available data points. To reduce the number of parameters the transition regions in $P(\gamma, \varphi)$ (red, cyan and brown regions in Fig. 94). were not included in the fit as they are not expected to give significant contributions on experimental time scales. The results of the least squares fit to the mica and hydrophilic graphite data are shown in Fig. 95 and the resulting a_{ij} are given in Appendix E. The results are supported by maximum entropy fits performed by G. Settanni that include the transition regions [346].

It is clear that adsorption on the mica surface can only be explained if correlations between the hinges are taken into account. Otherwise a strong peak at $\alpha = 180°$ is always present that is not observed in experiments. This is due to the fact that the dihedral values in the two protomers are shifted by π so that for $i = j$ the trans conformation is almost always present with similar values of the projected hinge angles Δ_1 and Δ_2 (see Fig. 93). This results in distributions $P_{ii}(\alpha)$ that are strongly peaked at $\alpha = 180°$. The uncorrelated model can not give more weight to off-diagonal elements that produce peaks at lower values of γ without increasing the peak at 180°.

The necessity for correlations that are not observed in solution simulations demonstrates that the adsorption on mica is due to specific interactions. These interactions are not included in the simple surface model where the distance of the domains to the surface is the only relevant parameter. Instead a better surface model should include an orientational dependence. In section 6.1 patchy electrostatic interactions were identified as an example of important interactions between fibrinogen and mica that depend on the orientation. Because of the missing flexible parts of the E-region, only a preferred orientation of the D-region can be inferred with a degree of certainty. Placing both D-regions in their preferred orientation on the surface would impose restrictions on the hinges. These would manifest as an apparent correlation of the two protomers.

Even with the correlations in place, the phenomenological model can only explain the dominant peak around 155°. The secondary peak around 115° is missing. In the maximum entropy fit this secondary peak can be reproduced better. This suggests that the regions neglected for the least square fit might be of more importance than initially assumed.

The results on hydrophilic graphite are consistent with an uncorrelated model. Because the statistics of the experiments on hydrophilic graphite are limited, those results should not be overinterpreted. In this case it could be concluded that the graphite surface acts simply as a sterical hindrance for fibrinogen. This is in accordance with the atomistic results in section 6.2.

For a better comparison more data of the fibrinogen bending angle distribution on various well characterized surfaces would be needed. Given the wealth of fibrinogen imaging studies these data should not be difficult to collect from experiments.

It would also be interesting to extend the model to multiple fibrinogen molecules that can interact on the surface to see what steric effects come into play when the surface coverage increases. This could be done using an RSA approach or MC techniques.

8.3 SUMMARY

In this chapter, a phenomenological model of fibrinogen was developed. This model is motivated by the atomistic simulations in Chapter 5 and captures only the main large scale feature observed in the atomistic simulations, the hinge bending. The model can be studied at a hard wall surface to understand conformational restrictions acting on fibrinogen at surfaces.

A comparison with the atomistic data in Chapter 6 was not attempted as the atomistic simulations are too short to access time scales on which the model is a good approximation for fibrinogen behavior. Using the angle distribution for the hinge obtained in the atomistic simulations it is possible to explain AFM experiments on mica and graphite surfaces. To this end, a fit of model parameters to experimental data is performed. To reproduce experimental results on a mica surface, a correlation between the hinges has to be introduced. This can be motivated by the earlier notion that fibrinogen adsorption on mica leads to assymetric adsorbed conformations.

Part IV

CONCLUSION

9

CONCLUSIONS AND OUTLOOK

In this work, the behavior of fibrinogen in solution and at inorganic surfaces has been investigated. A summary of the literature on fibrinogen has been done, giving a consistent picture of the biochemistry of fibrinogen, the self assembly process to a fibrin network, and how this has been studied experimentally and with modeling techniques.

The behavior of fibrinogen in solution and at two model surfaces was investigated using all-atom molecular dynamics simulations. These simulations showed a large scale motion enabled by a molecular hinge. The hinge behavior does not seem to depend on the dimerization or glycosilation state of fibrinogen. At surfaces this behavior is present but adsorption to the surface restricts or delays bending motions. Due to restrictions in the simulations, like the missing αC-regions or inaccuracies in the force field, an experimental verification of the large hinge bending motion is desirable. Some evidence for the existence of the hinge is provided by the highest resolution AFM images of adsorbed fibrinogen [17]. Higher resolution images, resolving the αC-region unequivocally, could establish the existence of the hinge further. A first such high resolution study has recently been published [64] and shows conformations with hinge bending. Additional evidence from FRET- or electron paramagnetic resonance techniques cold provide additional evidence for the existence of a hinge and its bending dynamics. To avoid labeling problems in the D-region a marker could be attached to an alanine-proline-arginine-proline peptide. This sequence binds exclusively in the β-pocket so that the position of the label is well known and no covalent modification of the globular domains would be necessary. The Fp-tether would be a prime target for the attachment of a marker in the E-region. Sum frequency generation (SFG) is another technique that would potentially allow a demonstration of the hinge. If fibrinogen were a stiff linear protein, no SFG signal would be observed as the dipole moments of the helices in the coiled-coil region cancel [112]. However a signal is observed. This has been attributed to a bending at the E-region and should be reexamined with a hinge model. In general, it would be interesting to revisit the interpretation of many experimental results that assume a rigid fibrinogen core or a bending in the E-region.

The basis for the molecular hinge are two proline residues that disrupt the α-helix of the γ-chain in the middle of the coiled-

coil region of fibrinogen. It has been found, that such a proline pattern is a well conserved feature of the fibrinogen sequences of all jawed vertebrates. Differences in the proline pattern are associated with major evolutionary changes in the clotting cascade. Disease mutations in human fibrinogen further suggest that the hinge is relevant for the physiological function of fibrinogen. Such a function could be the regulation of the exposure of plasmin binding sites. This is tentatively supported by the simulations. It is also possible that the flexibility provided by the hinge is essential for the correct formation of fibrin fibers and networks. Further experimental studies are needed to shed light on the physiological function of the hinge. Possibly mutation studies could be performed that exchange the prolines for alanines to suppress the formation of a hinge region. Such mutants can be expressed in mammalian cells [2]. If they can be isolated, activity studies of plasmin should be comparatively easy. Changes in the network structure could be accessed with the help of electron microscopy.

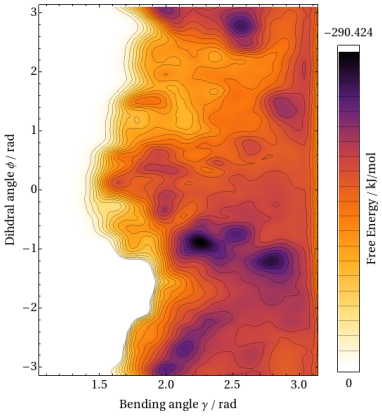

Figure 96: Preliminary free energy surface of the hinge determined in the Masters thesis of L. Ripka by well tempered metadynamics simulations [339].

The hinge was characterized by a bending and a dihedral angle. These angels are correlated and offer two easy bending directions. The simulation time, while long enough to identify the hinge, is not sufficient to obtain an equilibrated free energy landscape of the hinge. The bending itself is a rather fast phenomenon happing on a time scale of 20ns. However bending events are interspersed with long waiting times that can exceed 100ns. To obtain an equilibrium distribution of the hinge, extremely long trajectories would be necessary. Such a brute force approach is neither feasible nor necessary. Instead, more sophisticated sampling techniques, like metadynamics [347, 348], can be used. In this method, free energy minima are systematically filled with artificial barriers so that the sampling of new conformations is increased. Recording the placed barriers allows the reconstruction of the free energy surface. First studies of this have been performed in the Masters thesis of L. Ripka [339] and the preliminary results are shown in Fig. 96. So far the results of the metadynamics calculation are in agreement with the presented simulations. The metadynamics simulation still lack a sampling of highly bend conformations and further effort should be devoted to a reconstruction of the complete free energy surface of the hinge. These calculations could also include the secondary carbohydrate cluster in the hinge that could be reconstructed from the structure of the β-pocket glycans. Recent improvements in the force-fields for protein and sugars [165, 168] would lend additional accuracy to these calculations.

Many experiments are still necessary to establish the existence and function of the hinge beyond doubt. The simulations have nevertheless shed some light on speculations raised from experiments. One is that the hinge is much more flexible than previously thought. Hence the motion of the D-region with respect to the E-region is much greater than estimated from crystal struc-

tures. A bending at the E-region, as assumed in many imaging studies [17, 104, 112, 112, 143], also seems out of the question. Furthermore the simulations indicate that the flexibility is not contingent on the presence of the fibrinopeptides as was previously speculated [17]. The fibrinopeptides have been absent from the simulation but strong bending events were observed nevertheless.

Experiments have also speculated about the importance of the interaction of $\gamma 375$ with $\gamma 364$ for the formation of an open γ-pocket [319]. The small scale dynamics accessible in the atomistic simulation allow a characterization of the pockets. The central role played by $\gamma 375$ in the opening of the pocket was identified. The emerging picture is complicated by the fact that a simultaneous association of $\gamma 375$ with $\gamma 297$ and $\gamma 364$ is possible. Such an association leads to a closed pocket. Therefore, an association of $\gamma 375$ with $\gamma 364$ is not enough to guarantee an open pocket.

A possible allosteric connection of the two pockets was identified by the dependence of the γ-pocket on the glycosilation state of the β-pocket. Further simulations including the bound ligands would be helpful in establishing the existence and consequences of such a connection between the pockets. An extensive literature search did not bring any previous speculation about an allosteric effect to light. Consequently no attention has been paid to the investigation of this effect until now. The possible allosteric regulation of the two peptide binding pockets in the fibrinogen D-region is thus a fundamentally new behavior revealed in the atomistic simulations. As such, experimental investigations are needed to confirm its existence. In section 5.4.2, it was discussed that this could be done with fibrinogen variants that show mutations in the β-pocket. A new positive patch in the α-chain has also been identified that supports attachment of the carbohydrate clusters.

The data provided by the solution simulations allowed the development of a systematically coarse grained model of fibrinogen. In this model, the fibrinogen dimer is represented by 45 beads instead of nearly 31,000 atoms. The model recovers basic structural aspects like the hinge, the binding-pockets, and the Fp-tethers. The dynamics of a single dimer are reproduced by a heterogeneous elastic network. This network also revealed that the hinge forms a bottleneck, separating the dynamics in the central and outer regions of fibrinogen. In such models, the allosteric effect could be investigated better. In principle, the collected data would allow the construction of higher resolution models. Due to lack of time it was not investigated if this network reproduces the allosteric effect. If it does not, this effect should be included in some other way in the CG model. In principle, the collected data would allow the construction of higher resolution models. The higher resolution would provide a better network that represents more details of the communication pathway. It would also allow *in-silico* mutational studies. In such

studies, some of the fluctuations, e.g. in the β-pocket, could deliberately be changed before parametrization of the elastic network. The change in the resulting network can then be analyzed to see in which regions the most changes are concentrated. If no allosteric effect is present, the changes should be fairly localized, if it is present they should extend along the communication pathway to the γ-pocket.

A new method was presented to assign charges to the CG beads based on the atomistic simulations. This procedure captures the essential features of the fibrinogen charge structure better than a simple averaging. An attempt was made to parametrize the non-bonded interactions with a Lennard-Jones potential. The radius was obtained from the excluded water volume. For further studies the Lennard-Jones type interaction could be replaced by purely repulsive potentials as a first approximation. It is possible that even if the attractive part is included it might not be enough. The study of the γ-pocket revealed a transient helix forming in one of the loops. This likely effects the binding of the ligand and should be included in the model. A similar case has been identified in actin and methods exits that can incorporate such structural changes into the parametrization of coarse grained interactions [209]. The drawback of this method is that even more extensive atomistic simulations would be necessary.

The adsorption of fibrinogen was studied at different surfaces to provide insights for experiments performed under such conditions. Different adsorption behaviors were observed at the two inorganic surfaces studied. The hydrophobic graphite surface is very sticky. Many stable contacts are formed early on and little to no reorientation or desorption takes place. The adsorption is mainly driven by hydrophobic interactions. The strong interaction between protein and surface leads to deformations of the protein that can expose the binding sites more.

In contrasts to that, it was found that fibrinogen is only weakly adsorbed on mica. Only slight deformations of the binding sites occur. The most dramatic of which occurs in the β-pocket and is a consequence of specific interactions with the mica surface. Fibrinogen frequently left the mica surface and reoriented itself with respect to the surface. It was shown that the D- and E-region can reorient without strong sterical restriction imposed by the coiled-coil region. The simulations allowed the identification of likely favorable adsorption orientations of the two regions. The E-region will orient so that the Fp-tethers face the surface. The preferred orientation of the D-region has the P1-cleft facing the surface. The adsorption on mica is mainly driven by electrostatic forces and the subtle effect of charged patches on the fibrinogen surface has been discussed. Large parts of the flexible parts in the fibrinogen E-region as well as the αC-regions are missing in the simulations. In the future, it would be interesting to study how they affect the adsorption behavior. Especially the αC-regions are an important factor as they are known to in-

fluence adsorption and fibrin formation at surfaces [144, 145]. The charged patches on these regions likely influence their adsorption on mica [27] while it is known that changes in the dynamics of such flexible parts can support adhesion to hydrophobic materials [75, 80]. Also the inclusion of the complete second protomer would effect adsorption states and it would be interesting to see if a competition between the favorable orientations of both D-regions exists. Furthermore longer simulations are of interest to observe adsorption states that are closer to experiments. Such longer simulations would also allow the estimation of the surface diffusion of fibrinogen and its adsorption energy. To study structural transitions on the surface, steered MD simulations might be necessary that push the protein against the surface.

To compare implications from simulations to experimental adsorption studies a phenomenological model of fibrinogen was developed. This model captures only the hinge bending behavior and the geometry of fibrinogen. Using the distribution of the hinge bending and dihedral angle measured in the atomistic solution simulations the behavior of this model at hard surfaces was studied with a Monte Carlo approach. The results explain experimental results on mica and graphite. To reproduce the behavior at mica a correlation between the protomers needs to be established. This correlation highlights the fact that a preferred adsorption orientation exists on mica.

As is apparent from this summary, the results of this work open up many new avenues of research into fibrinogen behavior. First of all more materials surfaces could be studied to identify the different adsorption mechanisms caused by different surface chemistries. Among the more interesting of these is titanium as it is often used in medical implants. The problem with studying this surface is of course that metals are highly polarizable so that standard force fields can not reproduce the essential effects occuring during adsorption. A polarizable force field would be necessary. Alternatively, a hybrid simulation method like AdResS could be used to calculate the polarization on a quantum mechanical level at the site of contact formation.

Another important surface to study would be silica. It is one of the most widely used experimental surfaces and has the advantage that the effect of crystallinity could be explored. Experiments show that the degree of crystallinity is negatively correlated with protein adsorption [121] and simulations could explore the atomistic underpinnings of these effects. The challenge in simulating an amorphous silica surface is of course that a structure for a sufficiently large amorphous surface needs to be obtained first. A force field for different silica surface topologies has recently been published [241].

Other studies of silica surfaces have focused on peptide adsorption on nano-particles of different diameter [240]. With the

increasing interest in the applications of nano-particles for medical interventions, their interaction with fibrinogen becomes important. This means, the effect of a finite particle diameter also becomes of interest for fibrinogen adsorption. At this scale, the curvature of the surface is appreciable on the length scale of fibrinogen and a single fibrinogen molecule can connect two nano-particles [104]. These effects would require system sizes on the scale of several hundred nano-meter and can no longer be handled in atomistic simulations. Instead, use should be made of the coarse grained model developed in this work.

On the scale of nano-particles, hydrodynamics can no longer be neglected. Similarly, in medical devices both the surface chemistry and hydrodynamics influence adsorption behavior [76]. In principle, the explicit solvent in the atomistic simulation should include hydrodynamic phenomena. In CG simulation the solvent is replaced by an implicit model that should account for the correct hydrodynamics. The solvent is not the only thing that has to be treated carefully for hydrodynamic consistency. It is also known that the αC-regions need to be included to get the correct hydrodynamic behavior of fibrinogen [27]. On the CG level this could be done using a worm-like chain polymer model, an approach previously used in the description of fibrin fibers [253]. However other approaches exist to capture the hydrodynamic behavior of intrinsically disordered protein regions on a coarse grained level [215]. Similarly the complete Fp-tethers should be included in the CG model. Recently a model for these regions has been proposed [252].

Another aspect where hydrodynamics will become important is the behavior of fibrin fibers under blood flow. To study this, first of all the coarse grained model has to be developed far enough to reproduce reasonable fibrin structures. These could be compared to available experimental data [17]. In such a model, the self healing [349] of the fibers could be addressed. In CG simulations of fibrin fiber formation, it might be possible to determine the different steps and rate constants involved in fibrin formation. The determined values could then be used as input for a kinetic theory of fiber assembly that has been proposed over twenty years ago [247].

A successful CG model of the fibrin fiber should reproduce the strain hardening of the fiber. It might be necessary for this to perform further modeling (or hybrid simulations) of the ligand binding pockets to capture the catch-slip behavior they exhibit with the knobs [349]. To understand such a behavior free energy calculations of the knob-hole interactions under different conditions would be necessary. Once the correct fiber formation is established, it should be possible to look at the branching of fibers and thus the initial states in the formation of a fibrin network. Maybe it is even possible to develop a continuous description of the mechanical properties of fibrinogen fibers. The heterogeneous structure of these surfaces would likely yield a beam struc-

ture with fluctuating elastic constants. Methods to treat such systems exist [350].

To study layer formation on different surfaces it is not only important to develop a CG model of fibrinogen but also of the surface. A first step in this respect can been done by extending the proposed charge fitting procedure to include periodic boundary conditions. First steps in this direction have been presented here. More sophisticated techniques will be needed to capture chemical differences that go beyond the charge distribution. In such a CG model, the impact of different surface geometries and chemistries on the layer formation could be investigated. In the case of mica, it was already demonstrated in the atomistic simulations how the surface chemistry can lead to adsorbed conformations that block access to binding sites of fibrinogen. Hopefully, in the future the results of these studies can contribute to the design of materials that expose specific binding sites of fibrinogen and enhance the biocompatibility of these materials. The exposed binding sites of adsorbed fibrinogen can already today be investigated using functionalized AFM tips [130] so that comparisons to experiments should be possible.

The observed degree of flexibility makes it unlikely that a stable end-on adsorption conformation exists on surfaces. It is certainly plausible that the γ-pocket would be the only part of fibrinogen to make contact with the surface if a high concentration of fibrinogen is already adsorbed. However in this case the hinge bending would likely allow the rest of the protein to form a second layer of fibrinogen parallel to the surface. The processes leading to multilayer formation are also a problem that could be investigated using the coarse grained model of fibrinogen proposed in this work. These simulations as well as experimental studies [17, 46, 112] call multiple assumptions of Adamczycks RSA model for fibrinogen adsorption into question. The phenomenological model presented in this work is more in line with experimental findings and could be used as a replacement. Considering the many features that have been included in this description of fibrinogen, it would be interesting to see how the results change if the fibrinogen model is changed.

Another approach would be to further develop the simple Monte Carlo sampling scheme of adsorbed fibrinogen conformation into a full Monte Carlo simulation. In such a simulation, the competition between bending and crowding effects could be investigated. To realistically include the bending an accurate free energy surface of the hinge is necessary. This would allow a study of equilibrium coverage of a generic surface model as well as the study of adsorption/ desorption events leading to fluctuations in the equilibrium coverage that could be compared to, e.g. plasmon resonance experiments. To achieve more realistic surface interactions, it will be essential to include patchy interactions that give rise to the asymmetric adsorption events.

Concluding, it can be said that the presented simulations of fibrinogen highlight the importance of suspected features (hinge bending), identify new behaviors (allosteric effect) and elucidate experimental results (adsorption on mica). The developed models open up new avenues for further studies of fibrinogen behavior on multiple scales that could provide valuable insights in the future.

Part V

APPENDIX

A

FULL DATA FOR EVOLUTIONARY COMPARISONS

A.1 FIBRINOGEN SEQUENCE DATA

As described in section 4.2 the UniProt database was searched for fibrinogen sequences of different species. The evolutionary relationship between the identified species is shown in Fig. 97 and the UniProt identifiers are listed in Table 25 and Table 26. The hinge sequences for the different species are given in Table 27 and Table 28.

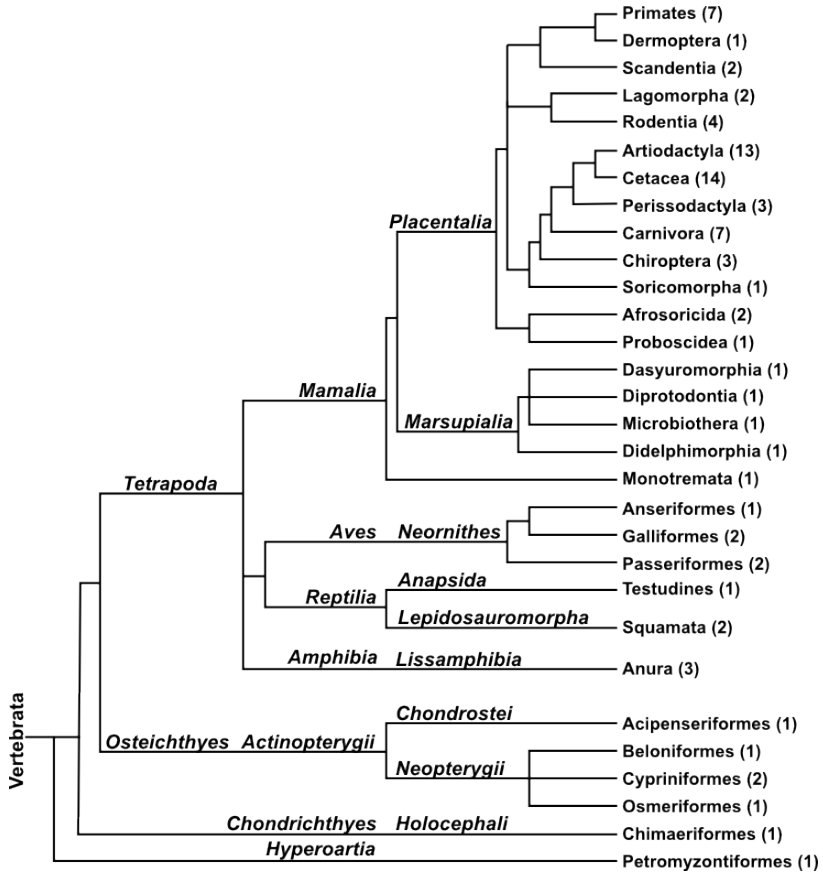

Figure 97: Taxonomic relationship between the orders for which fibrinogen sequences are available. The number behind the order siginifies the number of species in that order for which at least a partial γ-chain sequence is available. For details see Table 25 and Table 26.

Order	Species	Common Name	UniProt ID			
			α−chain	β−chain	γ−chain	
Primates	Homo Sapiens	Human	P02671	P02675	P02679	
Primates	Pan Troglodytes	Common Chimpanzee	H2QQB5	H2QQB4	H2RDH7	
Primates	Pongo Abelii	Sumatran Orangutan	H2PEK4	H2PEK2	H2PEK5	
Primates	Nomascus Leucogenys	Northern white-cheeked gibbon	G1R2L4	G1R2J2	G1R2S1	
Primates	Macaca Mulatta	Rhesus macaque	F6UZ60	F6UZ87	F6UYY8	
Primates	Callithrix Jacchus	Common marmoset	–	F7FU68	U3CZJ7	
Primates	Otolemur Garnettii	Northern greater galago	H0X4K1	H0X4J9	H0X4K3	
Dermoptera	Cynocephalus volans	Philippine flying lemur	–	–	Q6X870	*
Scandentia	Tupaia Chinensis	Chinease tree shrew	L8Y482	–	L8Y0J9	
Scandentia	Tupaia tana	Large treeshrew	–	–	Q6X865	*
Lagomorpha	Oryctolagus Cuniculus	Rabbit	G1T0X2	G1T0W8	G1TKX3	
Lagomorpha	Ochotona princeps	American pika	–	–	Q6X864	*
Rodentia	Mus Musculus	Mouse	Q99K47	Q8K0E8	Q8VCM7	
Rodentia	Rattus Norvegicus	Rat	P06399	P14480	P02680	
Rodentia	Heterocephalus Glaber	Naked mole rat	G5BML2	G5BML3	G5BML1	
Rodentia	Ictidomys Tridecemlineatus	Thirteen-lined ground squirrel	I3NCZ9	I3MEX9	I3LW67	
Artiodactyla	Bos Taurus	Cow	P02672	P02676	P12799	
Artiodactyla	Bos Mutus	Yak	L8I798	L8I9W3	L8I843	
Artiodactyla	Moschus sp.	Musk deer	–	–	A8IY89	*
Artiodactyla	Camelus Ferus	Wild Bactrian camel	–	S9XXW2	S9WDV3	
Artiodactyla	Camelus dromedarius	Dromedary	–	–	O02677	*
Artiodactyla	Lama guanicoe	Guanaco	–	–	A8IY99	*
Artiodactyla	Ovis dalli	Dall sheep	–	–	O12957	*
Artiodactyla	Sus scrofa	Pig	I3LQR9	F1RX37	F1RX35	
Artiodactyla	Tragulus napu	Greater mouse deer	–	–	O02690	*
Artiodactyla	Alces alces	Moose	–	–	O02672	*
Artiodactyla	Tayassu tajacu	Collared peccary	–	–	O12959	*
Artiodactyla	Giraffa camelopardalis	Giraffe	–	–	O02683	*
Artiodactyla	Hexaprotodon liberiensis	Pygmy hippopotamus	–	–	O12954	*
Cetacea	Lagenorhynchus obscurus	Dusky dolphin	–	–	A8IY92	*
Cetacea	Inia geoffrensis	Amazon river dolphin	–	–	A8IY96	*
Cetacea	Grampus griseus	Risso's dolphin	–	–	F5CIN4	*
Cetacea	Phocoenoides phocoena	Harbor porpoise	–	–	F5CIN5	*
Cetacea	Phocoenoides dalli	Dall's porpoise	–	–	F5CIN6	*
Cetacea	Mesoplodon peruvianus	Pygmy beaked whale	–	–	F5CIN7	*
Cetacea	Tasmacetus shepherdi	Shepherd's beaked whale	–	–	F5CIN8	*
Cetacea	Delphinapterus leucas	Beluga whale	–	–	O02681	*
Cetacea	Kogia breviceps	Pygmy sperm whale	–	–	F5CIN9	*
Cetacea	Physeter catodon	Sperm whale	–	–	O02687	*
Cetacea	Megaptera novaeangliae	Humpback whale	–	–	F5CIP0	*
Cetacea	Eschrichtius gibbosus	California gray whale	–	–	F5CIP1	*
Cetacea	Caperea marginata	Pigmy right whale	–	–	F5CIP2	*
Cetacea	Balaenoptera physalus	Finback whale	–	–	O02673	*
Perissodactyla	Equus Caballus	Horse	F6RUZ6	F6PH38	F6W2Y1	
Perissodactyla	Equus caballus przewalskii	Przewalski's horse	–	–	O02682	*
Perissodactyla	Tapirus indicus	Asiatic tapir	–	–	O02689	*

Table 25: UniProt ID for the used fibrinogen sequences from primates to odd-toed ungulates, evolutionary relations are depicted in figure 97. Sequences designated with * contain only a fragement of the coiled coil region. Sequences designated with † contain the complete coiled-coil region but an incomplete globular domain.

Order	Species	Common Name	UniProt ID			
			α−chain	β−chain	γ−chain	
Carnivora	Ailuropoda Melanoleuc	Giant Panda	G1MJI0	G1MJH9	G1MJI1	
Carnivora	Canis Familiaris	Dog	F1PBL4	F1PW65	F1P8G0	
Carnivora	Canis latrans	Coyote	–	–	O02680	*
Carnivora	Mustela Putorius Furo	European domestic ferret	M3Y2M4	M1EME9	M1ERI8	†
Carnivora	Neovison Vison	American mink	–	U6DW14	U6CRK3	
Carnivora	Felis Catus	Cat	M3W022	M3WII3	M3WN28	
Carnivora	Crocuta crocuta	Spotted hyena	–	–	O02676	*
Chiroptera	Myotis Davidii	David's myotis	L5LHG7	L5LGL2	L5LH50	†
Chiroptera	Myotis Lucifugus	Little brown bat	G1P500	G1P9V0	G1P505	
Chiroptera	Nyctimene albiventer	Common tube-nosed fruit bat	–	–	Q6X872	*
Soricomorpha	Scalopus aquaticus	Eastern mole	–	–	Q6X871	*
Afrosoricida	Amblysomus hottentotus	Hottentot golden mole	–	–	Q6X868	*
Afrosoricida	Tenrec ecaudatus	Common tenrec	–	–	Q6X869	*
Proboscidea	Loxodonta Africana	African bush elephant	–	G3TIT4	G3TIU1	
Dasyuromorphia	Sarcophilus Harrisii	Tasmanian devil	G3WMC2	G3WMN1	G3WLY0	
Diprotodontia	Phascolarctos cinereus	Koala	–	–	Q6X867	*
Microbiotheria	Dromiciops gliroides	Monito del Monte	–	–	Q6X866	*
Didelphimorphia	Monodelphis Domestica	Gray short-tailed opossum	F7CJ60	F7FX25	F7CIN4	
Monotremata	Ornithorhynchus anatinus	Platypus	F7CSN1	F7CST7	F7CSA4	
Anseriformes	Anas Platyrhynchos	Duck	U3I9V9	–	R0LIW0	
Galliformes	Gallus Gallus	Chicken	P14448	Q02020	O93568	
Galliformes	Meleagris gallopavo	Turkey	G1MRL1	G3UTA1	G1MRP5	
Passeriformes	Taeniopygia guttata	Zebra finch	–	H0Z4C6	H0Z4E6	
Passeriformes	Ficedula albicollis	Collared flycatcher	U3KF97	U3KF81	U3KFB4	
Testudines	Pelodiscus sinensis	Chinese softshell turtle	K7FND3	K7FPL9	K7FHU9	
Squamata	Anolis carolinensis	Carolina anole	G1KRE2	G1KRF4	G1KRE0	
Squamata	Ophiophagus hannah	King cobra	–	V8P046	V8P0Z0	
Anura	Xenopus Laevis	African clawed frog	Q801E3	Q91589	P17634	
Anura	Xenopus Tropicalis	Western clawed frog	F6SP35	B0JZ09	F6SPR7	
Anura	Hymenochirus Curtipes	Western dwarf clawed frog	G5DXW7	–	G5E070	†
Acipenseriformes	Acipenser sinensis	Chinese Sturgeon	–	–	D6R7B4	
Beloniformes	Oryzias melastigma	Indian Medaka	–	Q6NYE1	I1SRK4	†
Cypriniformes	Hypophthalmichthys molitrix	Silver Carp	–	–	D6R707	
Cypriniformes	Danio Rerio	Zebrafish	Q6DHS2	Q6NYE1	A8E5P0	†
Osmeriformes	Plecoglossus Altivelis	Ayu	–	F8K8M7	F8K8M8	
Chimaeriformes	Callorhinchus milii	Australian ghostshark	–	K4FUN8	K4FTB5	
Petromyzontiformes	Petromyzon Marinus	Sea lamprey	P02674	P02678	P04115	

Table 26: UniProt ID for the used fibrinogen sequences from carnivorans to lampreys, evolutionary relations are depicted in figure 97. Sequences designated with * contain only a fragement of the coiled coil region. Sequences designated with † contain the complete coiled-coil region but an incomplete globular domain.

Human	...TYNPDESSKPNMIDAAT...
Common Chimpanzee	...TYNPDESSKPNMIDAAT...
Sumatran Orangutan	...SYNPDEPSKPNTIDSAT...
Northern white-cheeked gibbon	...SYNPDESSKPNTIDSAT...
Rhesus macaque	...SYNPDESSKPNTIDTAT...
Common marmoset	...SYKPDEPSKPNMIEGAT...
Northern greater galago	...SYNPDQPSKPNMIEGAT...
Philippine flying lemur	...DYNPGEPPKQSVTEGAT...
Chinease tree shrew	...VYNPDGPSKPNMIESAT...
Large treeshrew	...IYNPDGPSKPNVIESAT...
Rabbit	...SYDTAEPSVPNKIDSAT...
American pika	...SYNPNEPSAPNKIESAT...
Mouse	...YYNPDQPPKPGMIDSAT...
Rat	...YYNPDQPPKPGMIEGAT...
Naked mole rat	...SYDPHQPAKPNAIETAT...
Thirteen-lined ground squirrel	...SYNPDQPKKPNMIETAT...
Cow	...SYNPDQPSKPNNIESAT...
Yak	...SYNPDQPSKPNNIESAT...
Musk deer	...SYNPDEPSKPNKIEGAT...
Wild Bactrian camel	...SYNPAEPSKPSRIESAT...
Dromedary	...SYNPAEPSKPSRIESAT...
Guanaco	...SYNPAEPSKPNRIESAT...
Dall sheep	...SYNPDEPSKPSNIESAT...
Pig	...SYNPEDLSKPDRIQSAT...
Greater mouse deer	...SYNPDEASKPNKIESAT...
Moose	...SYNPDEPSKPNNIESAT...
Collared peccary	...SYNPDXPSKPDRIQSAT...
Giraffe	...SYNPDEPSKPSHIESAT...
Pygmy hippopotamus	...SYNPDEPEKPSRIESAT...
Dusky dolphin	...SYHSDGPAKPNGIESAT...
Amazon river dolphin	...SYRSDGSAKPSGMESAT...
Risso's dolphin	...SYHSDGPAKPNGIESAT...
Harbor porpoise	...SYHSDGPAXPSGIESAT...
Dall's porpoise	...SYHSDGPAKPSGIESAT...
Pygmy beaked whale	...SYHSDGSAKPNGIESAT...
Shepherd's beaked whale	...SYRSDGSAKPNGIESAT...
Beluga whale	...SYRSDGPAKPNGIESAT...
Pygmy sperm whale	...SYRSDGPGKPSGIDSAT...
Sperm whale	...SYRSDGPAKPSGIESAT...
Humpback whale	...SYXSDGPAKPNGIDSAT...
California gray whale	...SYRSDGPAKPNGIDSAT...
Pigmy right whale	...SYRSDGPAKPNGIESAT...
Finback whale	...SYRSDGPAXPNGIDSAT...
Horse	...SYNPDEPPKTGRIDAAT...
Przewalski's horse	...SYNPDEPPKTGRIDAAT...
Asiatic tapir	...RYKPDEPTKPGGIDSAT...

Table 27: Amino acid sequence of the fibrinogen γ-chain in the hinge region for all available species from primates to odd-toed ungulates.

Giant Panda	...TYNPSEPPKPNRIVSAT...
Dog	...TYNPDEPPKPNRVVGAT...
Coyote	...TYNPDEPPKPNRVVGAT...
European domestic ferret	...TYNPNEPPKPNRIVSAT...
American mink	...TYNPNEPPKPNRIVSAT...
Cat	...NYNPNEPPKPNRIVSAT...
Spotted hyena	...SYNPNEPPKPNTIVSAT...
David's myotis	...FYSTDGLPKSDRIEASI...
Little brown bat	...VYNPDGLPKSDKIQAAI...
Common tube-nosed fruit bat	...NYDPNEPPKPNKIDSAT...
Eastern mole	...SYLPEDPARPNRVELAT...
Hottentot golden mole	...NYNPDEPTRPNMIQGAT...
Common tenrec	...SYNPDEPPRPNMIEGAT...
African bush elephant	...SYTPEEPSKPGAIQRAT...
Tasmanian devil	...GQGSD-PRPQNVINDAM...
Koala	...GRNSD-ARPQNVINDVT...
Monito del Monte	...GQTPD-PRPQNVIGDVT...
Gray short-tailed opossum	...GQAPD-PRPQNVINDVT...
Platypus	...G-GPDRSREPQNRQGVT...
Duck	...LYPPDKQQQQQSVEGFT...
Chicken	...IYPSEKQTLPQSIEQLT...
Turkey	...IYPAEKQTLPQSVESLT...
Zebra finch	...LYPSEKQTLPNSVDDFT...
Collared flycatcher	...LYPSDKQTLPNSIDDFT...
Chinese softshell turtle	...LSPPEK-TQPSTIDGFT...
Carolina anole	...VHTPNPSTQPNVIPGYT...
King cobra	...VHVSHPAAQSNVIPGYT...
African clawed frog	...SGKKPATSPQTAIDPMT...
Western clawed frog	...SGQKPVTST-PTIDPVT...
Western dwarf clawed frog	...SGKKPATSQ-QTIDPMT...
Chinese Sturgeon	...SQTQA--QKSSP-DLYI...
Indian Medaka	...SATSA--QKSSQPEPLF...
Silver Carp	...SETQA--QKQAPADTYI...
Zebrafish	...SEAQA--QKQSP-DTYI...
Ayu	...SVTSA--QKSSSPDQYF...
Australian ghostshark	...SNVVIK---KQLPDTFI...
Sea lamprey	...DVRITRDEAQIIKDSGQ...

Table 28: Amino acid sequence of the fibrinogen γ-chain in the hinge region for all available species from carnivorans to lampreys.

A.2 HELICITY

The following figures show the helicity of the three fibrinogen chains in the coiled-coil region of different species. Only species were considered for which the complete sequence of the coiled-coil region in all three chains was available. The residue numbers are measured from the first cysteine in the disulfide ring.

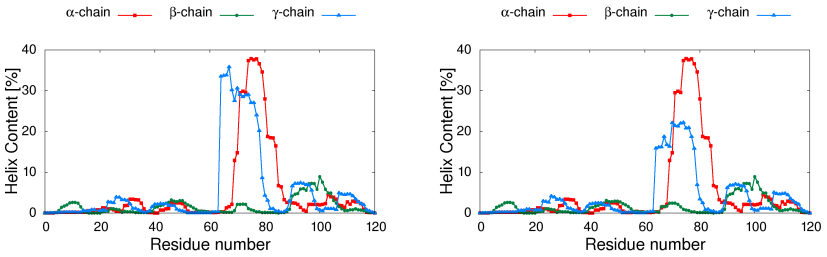

Figure 98: Left: Human, Right: Chimpanzee

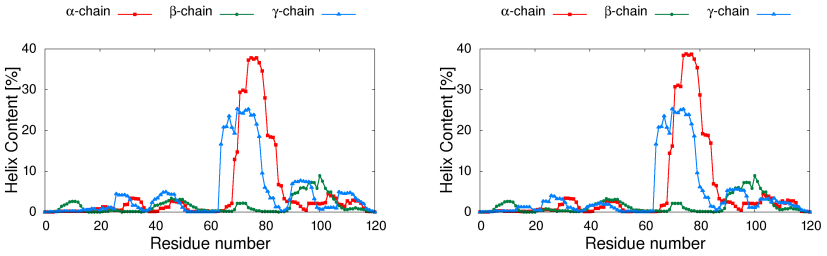

Figure 99: Left: Orangutan, Right: Gibbon

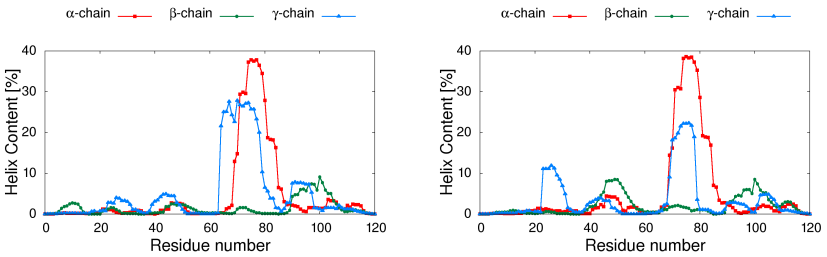

Figure 100: Left: Macaque, Right: Galago

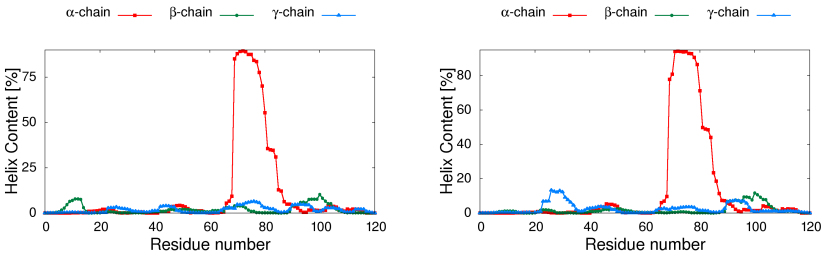

Figure 101: Left: Rabbit, Right: Mouse

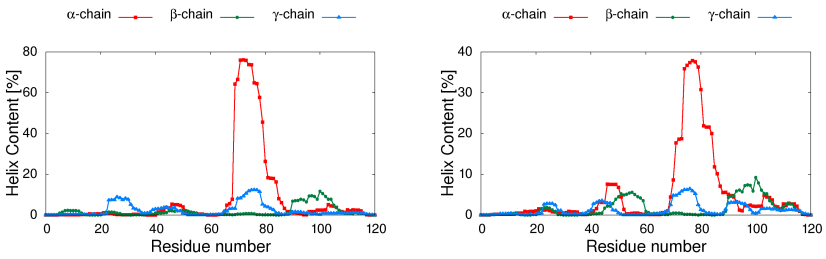

Figure 102: Left: Rat, Right: Naked mole rat

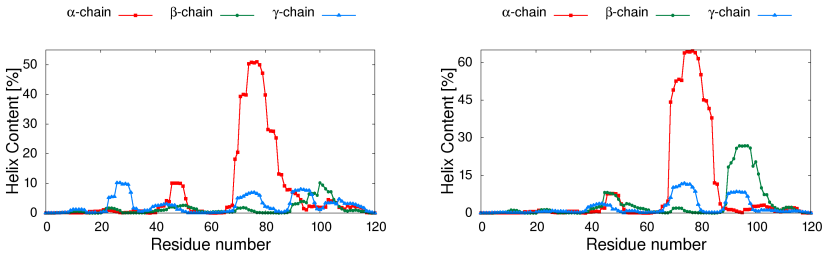

Figure 103: Left: Ground squirrel, Right: Pig

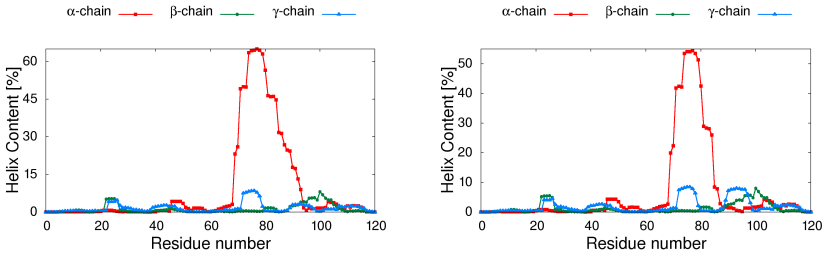

Figure 104: Left: Cow, Right: Yak

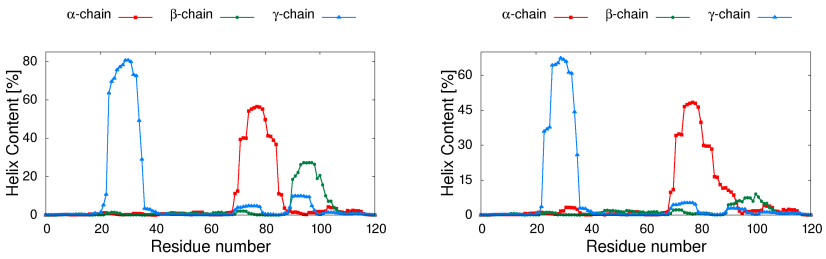

Figure 105: Left: Dog, Right: Panda

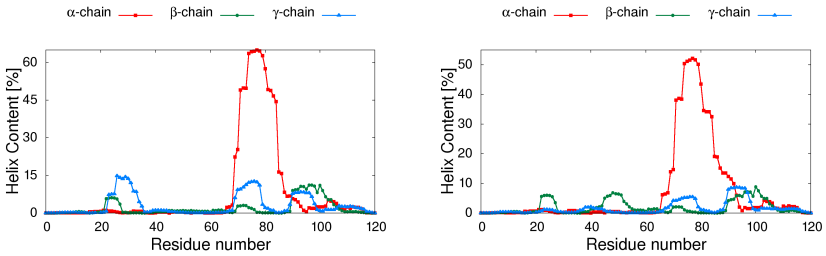

Figure 106: Left: Cat, Right: Ferret

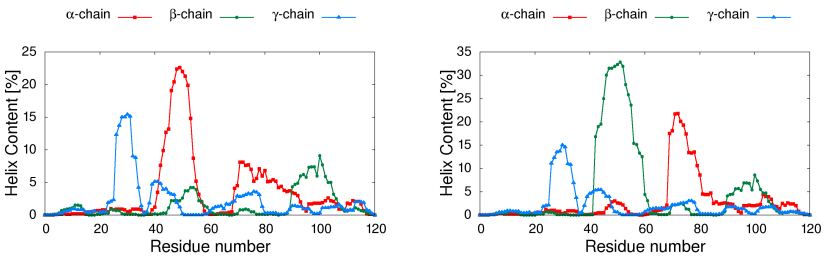

Figure 107: Left: David's Myotis, Right: Little brown bat

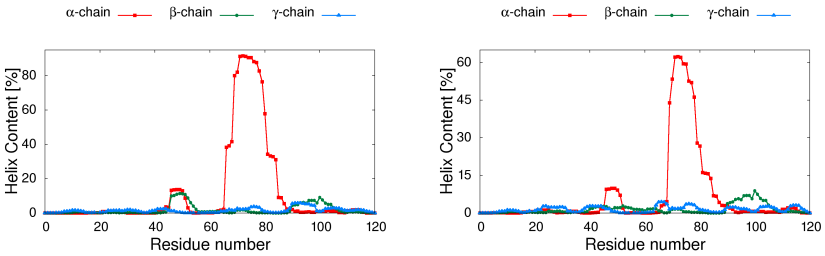

Figure 108: Left: Opossum, Right: Tasmanian Devil

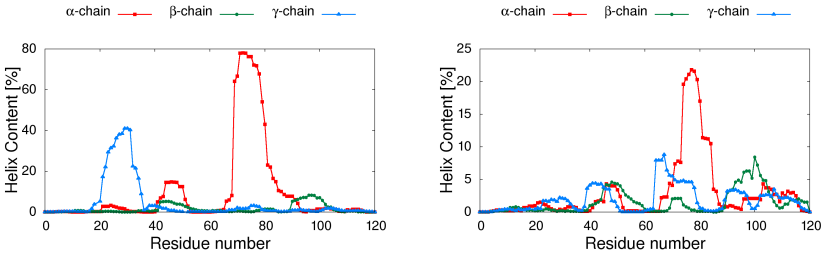

Figure 109: Left: Plattypus, Right: Horse

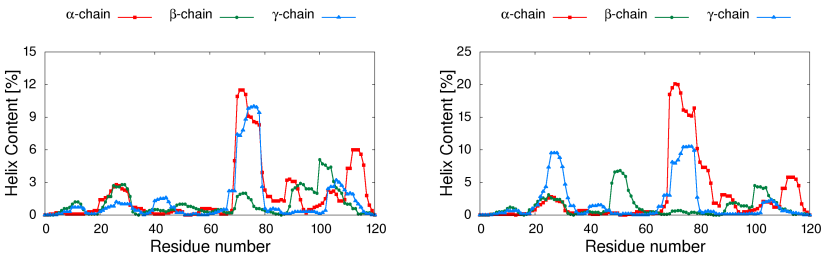

Figure 110: Left: Chicken, Right: Turkey

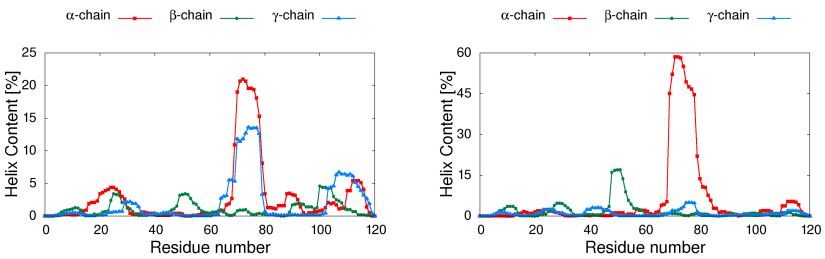

Figure 111: Left: Flycatcher, Right: Anole

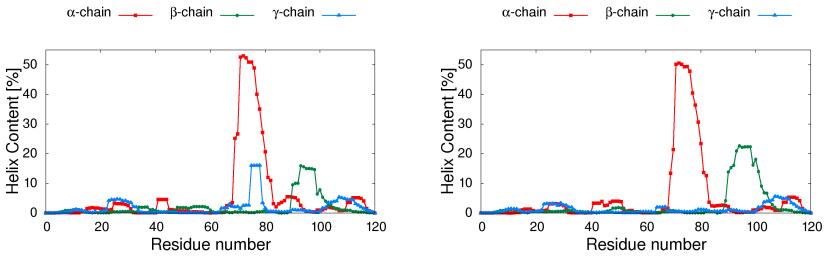

Figure 112: Left: Western clawed frog, Right: African clawed frog

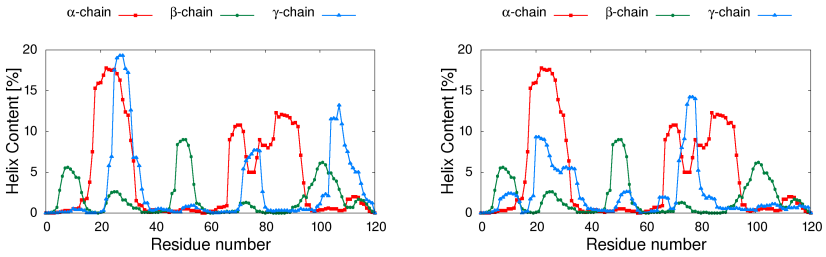

Figure 113: Left: Indian medaka, Right: Zebrafish

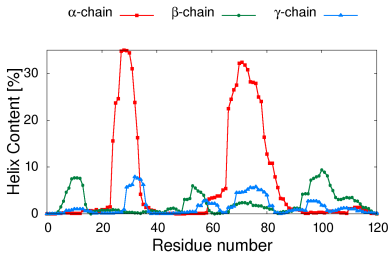

Figure 114: Lamprey

A.3 DISORDER: COILS

The following figures show the probability to form coils of the three fibrinogen chains in the coiled-coil region of different species. Only species were considered for which the complete sequence of the coiled-coil region in all three chains was available. The residue numbers are measured from the first cysteine in the disulfide ring. The gray line represents the threshold of p=0.516 above which residues are considered to be in a coil [292].

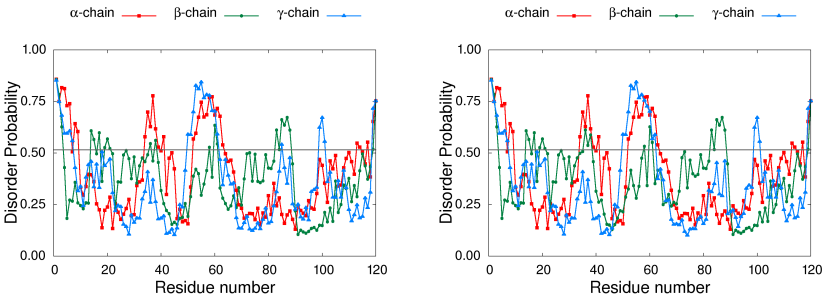

Figure 115: Left: Human, Right: Chimpanzee

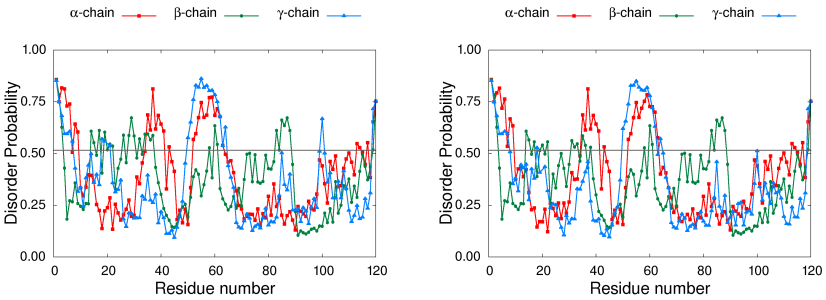

Figure 116: Left: Orangutan, Right: Gibbon

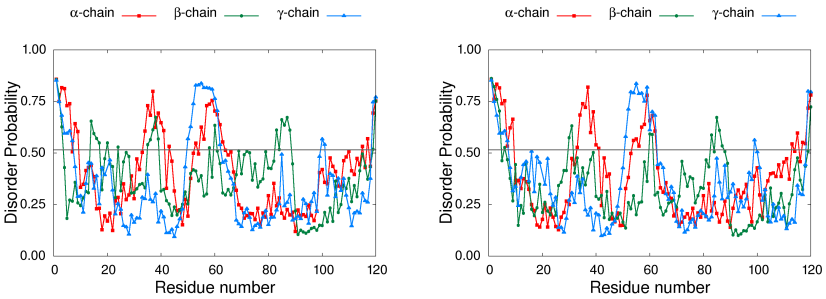

Figure 117: Left: Macaque, Right: Galago

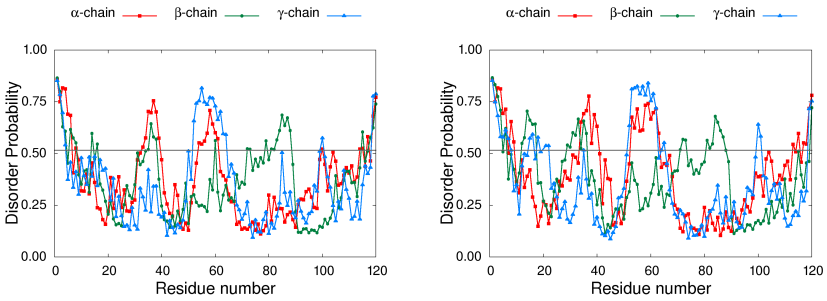

Figure 118: Left: Rabbit, Right: Mouse

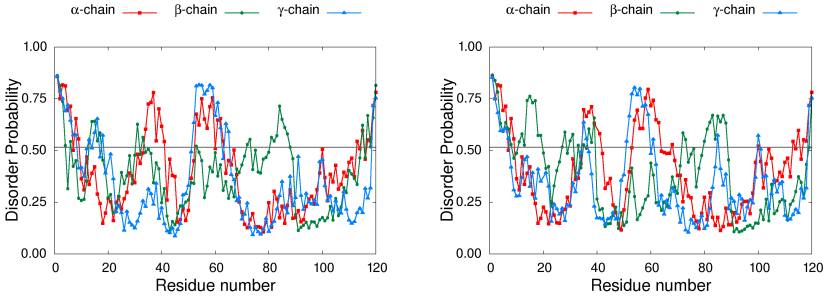

Figure 119: Left: Rat, Right: Naked mole rat

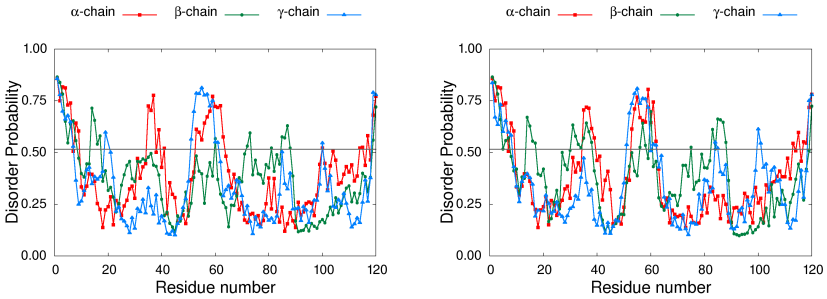

Figure 120: Left: Ground squirrel, Right: Pig

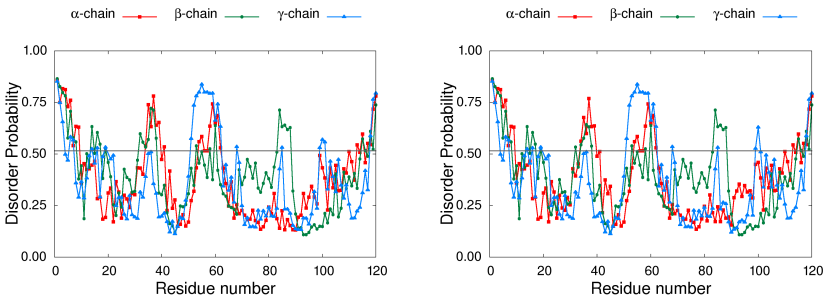

Figure 121: Left: Cow, Right: Yak

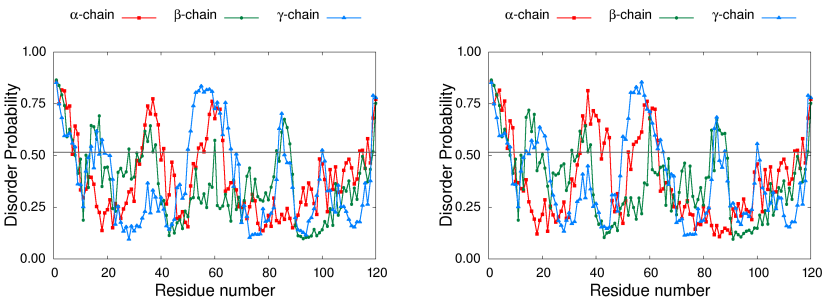

Figure 122: Left: Dog, Right: Panda

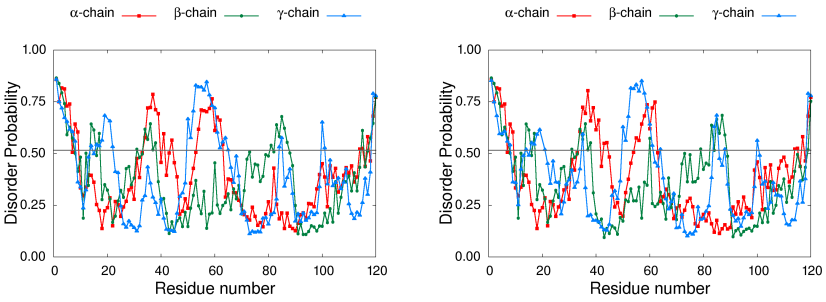

Figure 123: Left: Cat, Right: Ferret

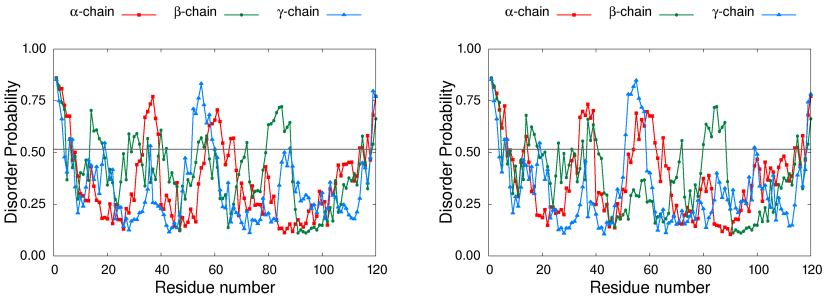

Figure 124: Left: David's Myotis, Right: Little brown bat

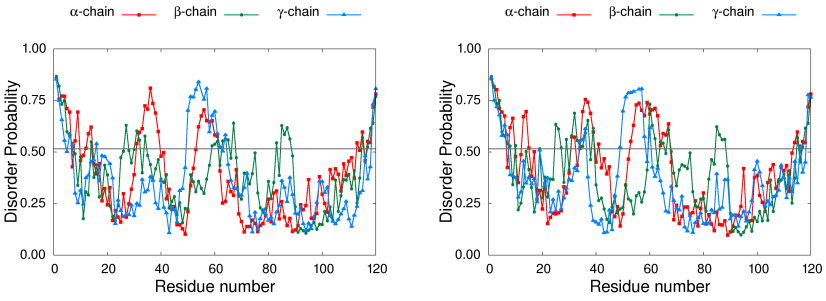

Figure 125: Left: Opossum, Right: Tasmanian Devil

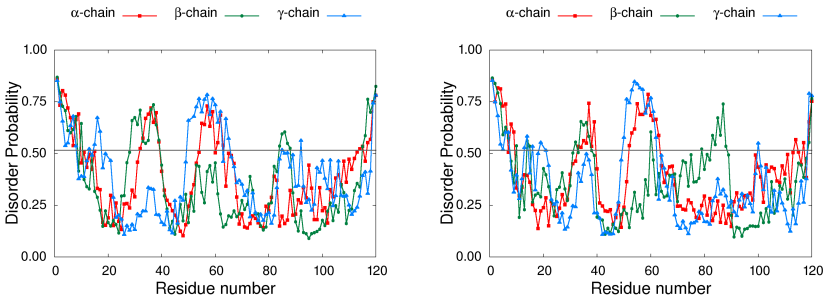

Figure 126: Left: Plattypus, Right: Horse

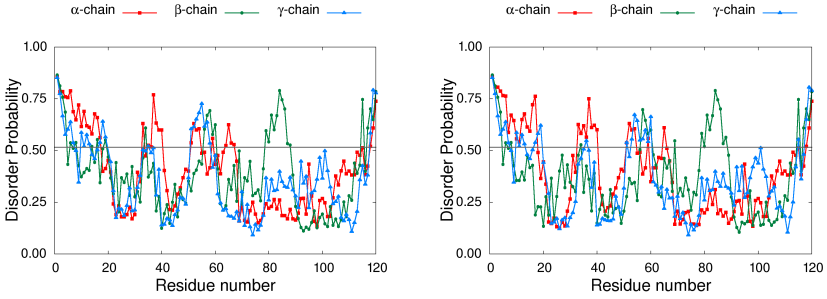

Figure 127: Left: Chicken, Right: Turkey

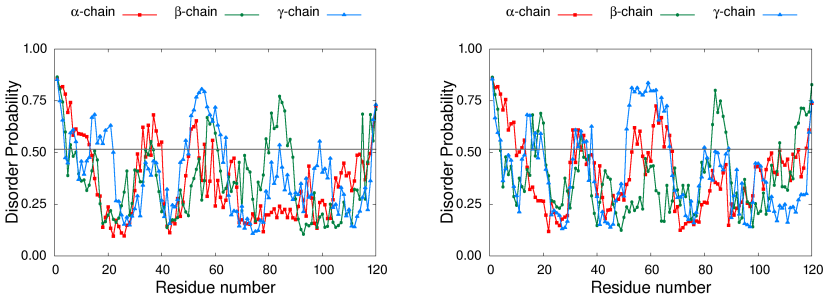

Figure 128: Left: Flycatcher, Right: Anole

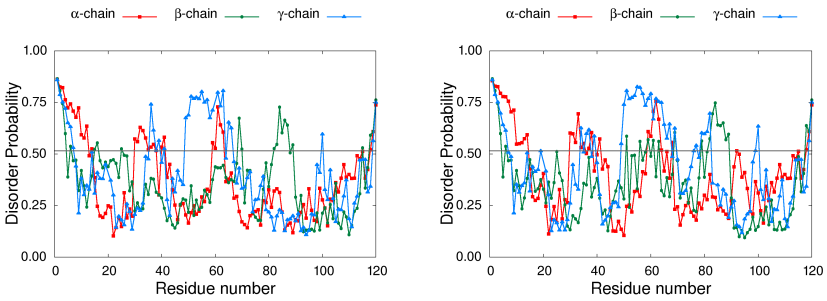

Figure 129: Left: Western clawed frog, Right: African clawed frog

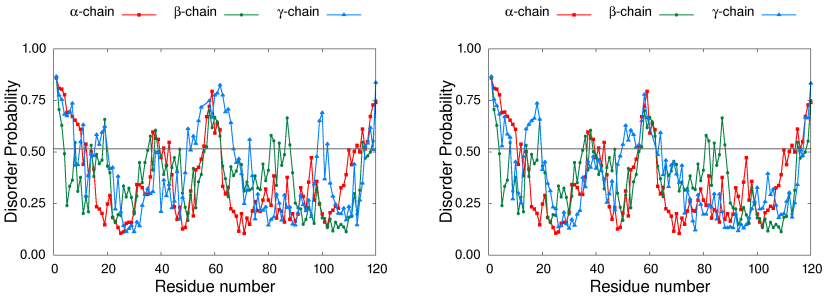

Figure 130: Left: Indian medaka, Right: Zebrafish

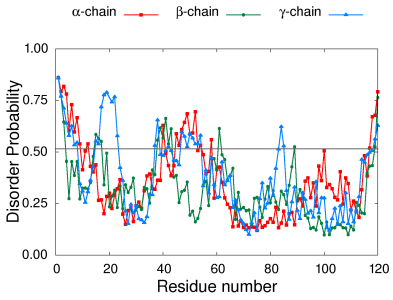

Figure 131: Lamprey

A.4 DISORDER: HOT COILS

The following figures show the probability to form hot coils of the three fibrinogen chains in the coiled-coil region of different species. Only species were considered for which the complete sequence of the coiled-coil region in all three chains was available. The residue numbers are measured from the first cysteine in the disulfide ring. The gray line represents the threshold of p=0.1204 above which residues are considered to be in a hot coil [292].

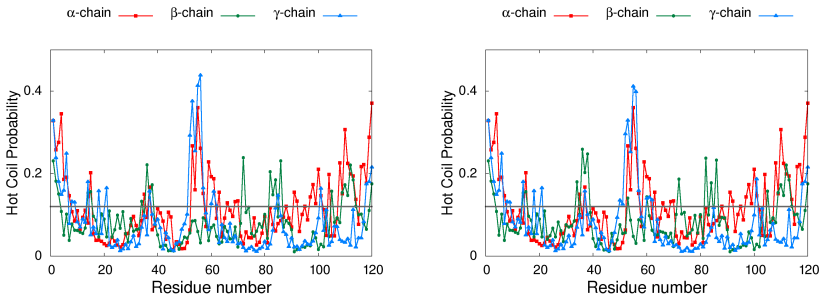

Figure 132: Left: Human, Right: Chimpanzee

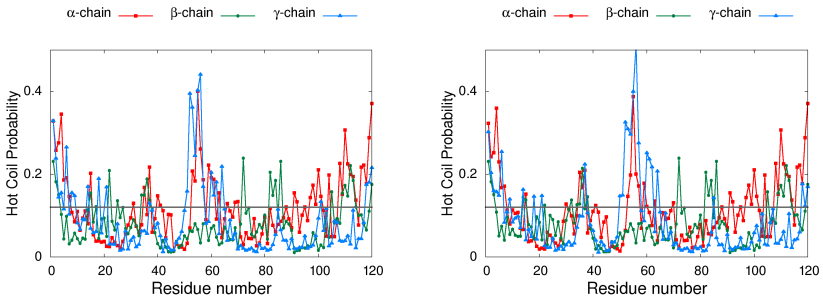

Figure 133: Left: Orangutan, Right: Gibbon

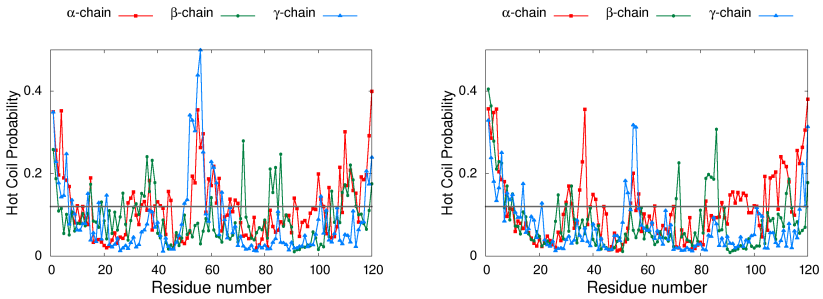

Figure 134: Left: Macaque, Right: Galago

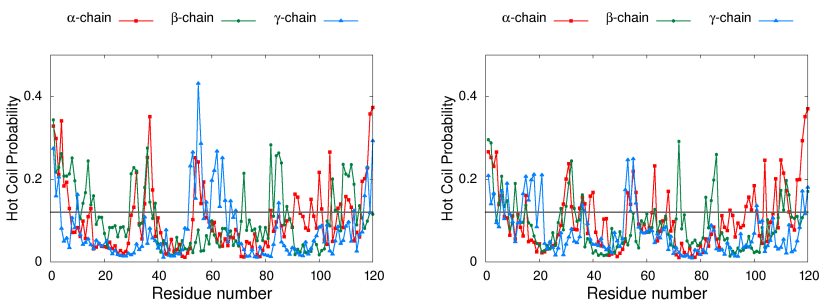

Figure 135: Left: Rabbit, Right: Mouse

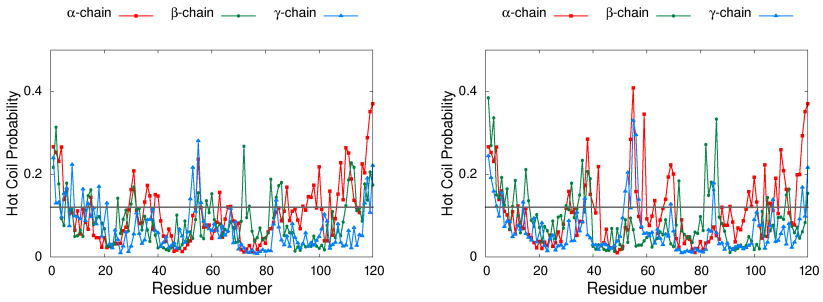

Figure 136: Left: Rat, Right: Naked mole rat

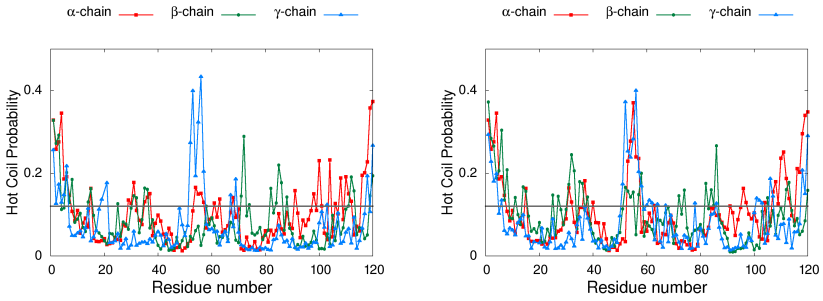

Figure 137: Left: Ground squirrel, Right: Pig

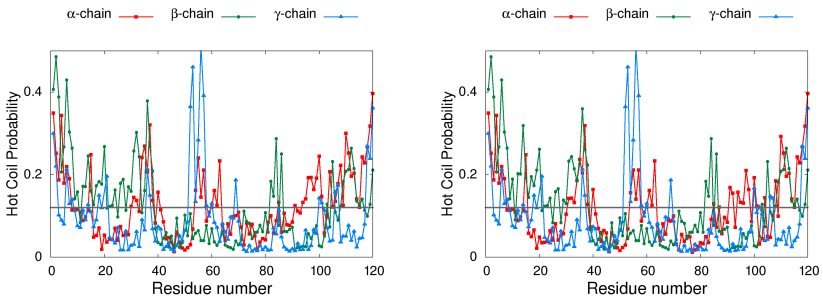

Figure 138: Left: Cow, Right: Yak

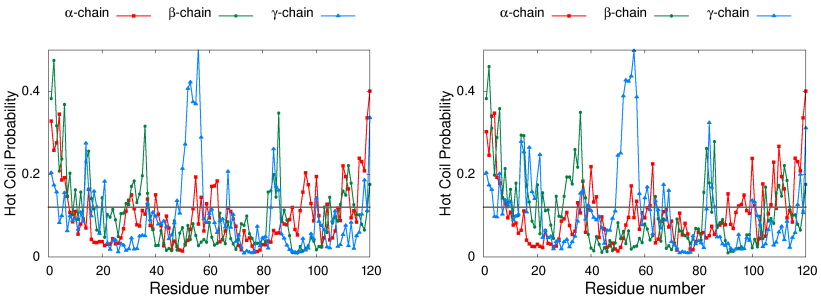

Figure 139: Left: Dog, Right: Panda

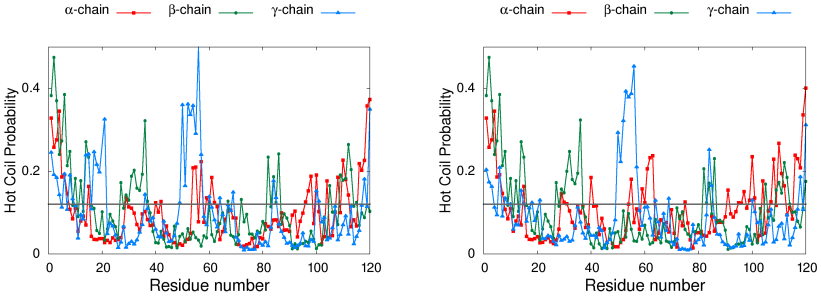

Figure 140: Left: Cat, Right: Ferret

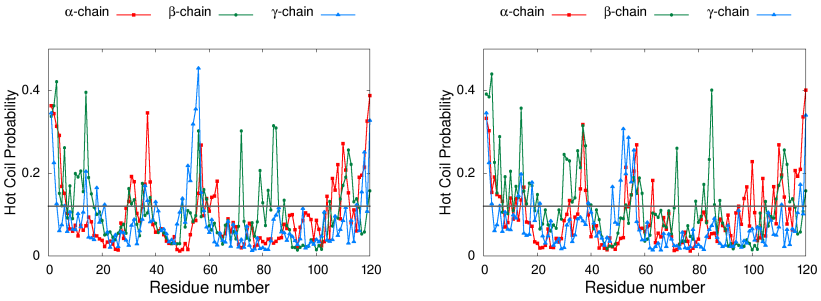

Figure 141: Left: David's Myotis, Right: Little brown bat

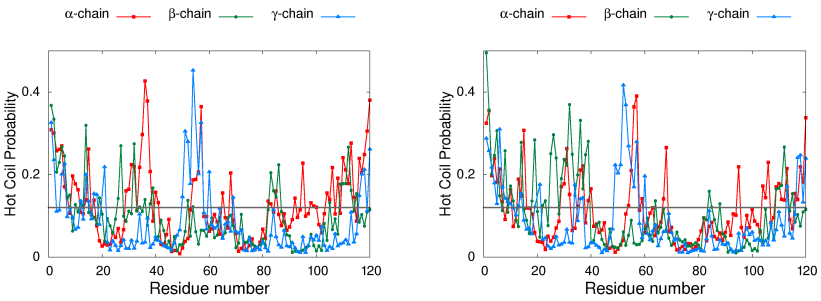

Figure 142: Left: Opossum, Right: Tasmanian Devil

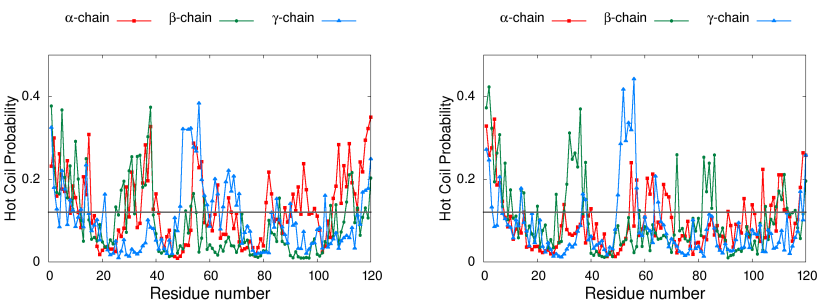

Figure 143: Left: Plattypus, Right: Horse

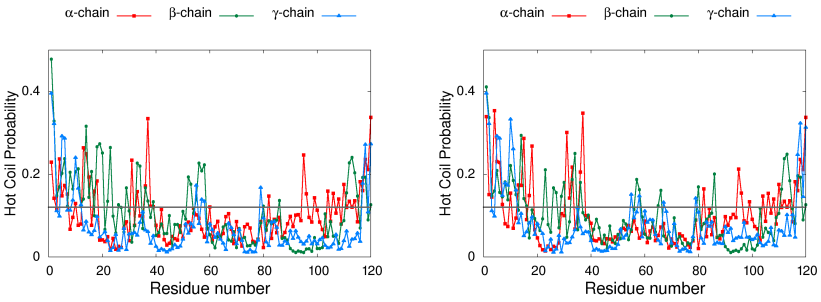

Figure 144: Left: Chicken, Right: Turkey

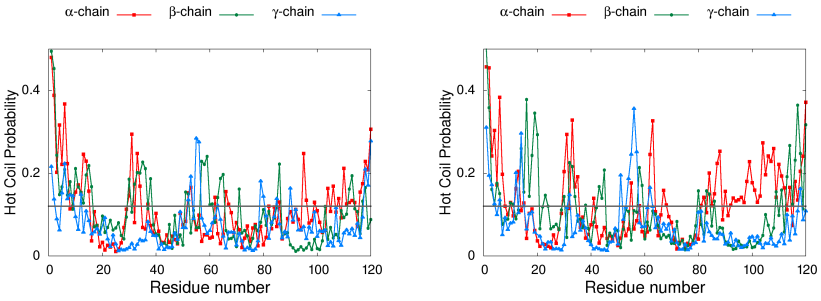

Figure 145: Left: Flycatcher, Right: Anole

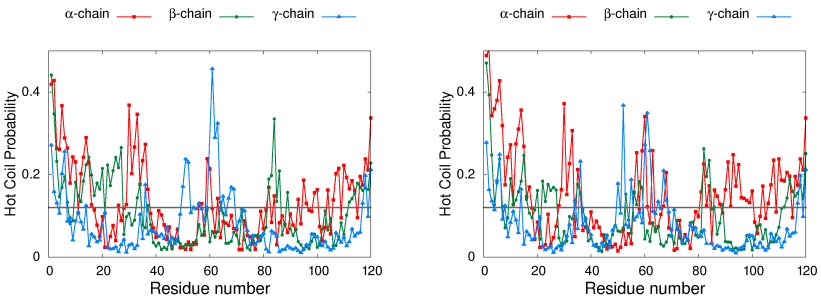

Figure 146: Left: Western clawed frog, Right: African clawed frog

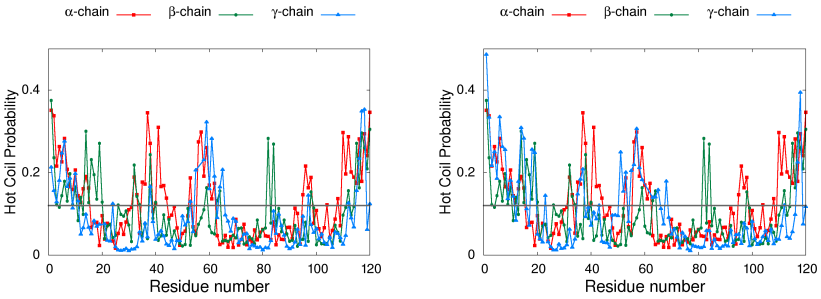

Figure 147: Left: Indian medaka, Right: Zebrafish

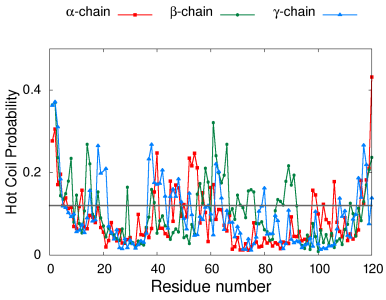

Figure 148: Lamprey

B

VISUALIZATION OF ADSORPTION EVENTS

B.1 MICA

The following figure shows the different conformations of fibrinogen adopted during the adsorption event described in section 6.1.2.

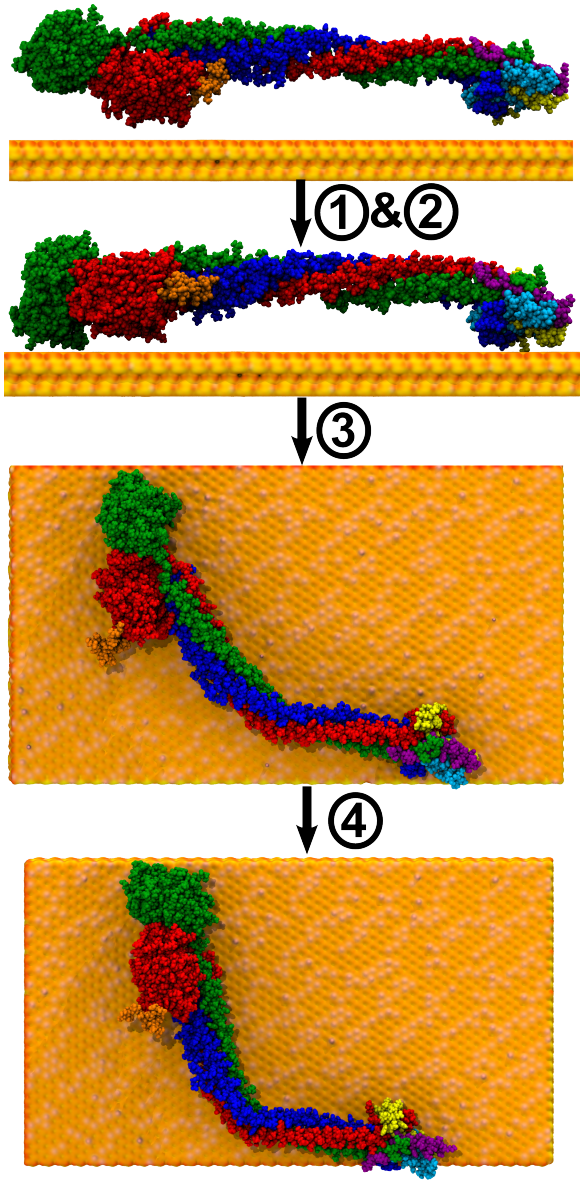

Figure 150: Snapshots of the fibrinogen adsorption process quantified in Fig. 149.

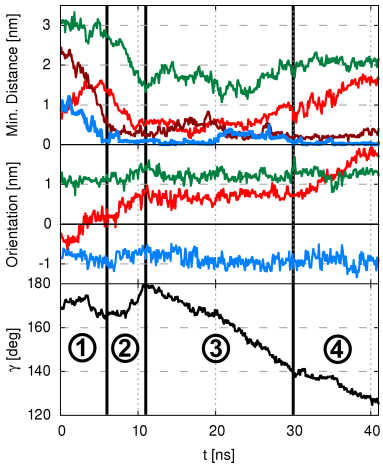

Figure 149: Quantities characterizing an adsorption event of fibrinogen on mica. CQuantities for the hinge are shown in green, for the E-region in blue and for the D-region in red (γC-domain in dark red).

B.2 GRAPHITE

The following figure shows the different conformations of fibrinogen adopted during the adsorption event described in section 6.1.2. Figure 152 shows the flattening of domains during the adsorption to graphite.

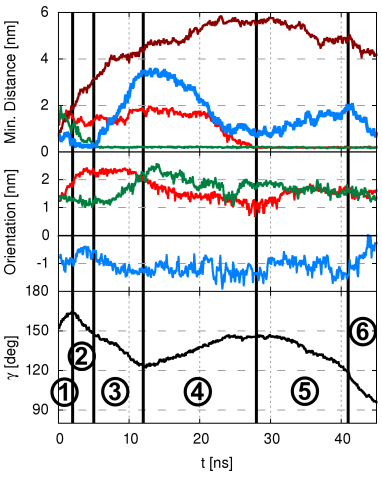

Figure 151: Quantities characterizing adsorption of fibrinogen on graphite. Quantities for the hinge are shown in green, for the E-region in blue and for the D-region in red (γC-domain in dark red).

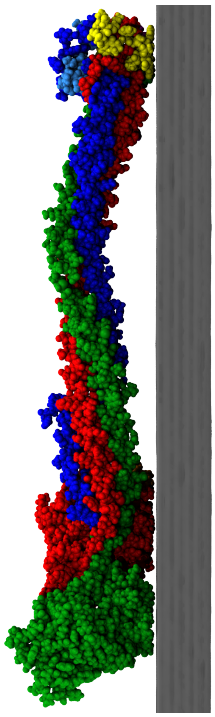

Figure 152: Flattening of the fibrinogen domains as a first step in denaturing during adsorption.).

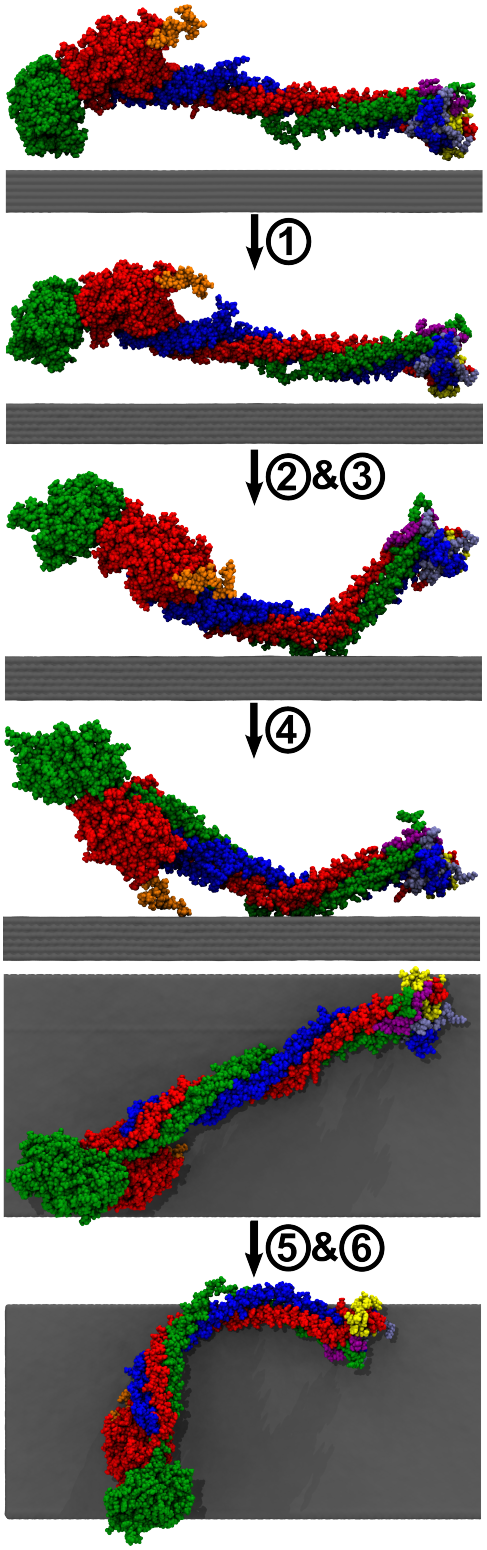

Figure 153: Snapshots of the fibrinogen adsorption process quantified in Fig. 72.

C

AN EFFICIENT ALGORITHM TO COMPARE NUMBERINGS OF SETS

During space based EDCG it is important to be able to compare different outcomes of the stochastic minimization. The outcome of such a minimization is a set of sets numbered from 1 through n. Each set contains the residues belonging to a particular CG site. During minimization the composition of the sets changes in a stochastic way so that the numbering can not be predicted. To compare outcomes from different minimizations, it is necessary to find a way to consistently number the sets of residues, i.e. construct permutations of the numbered sets. Previously an exhaustive search had been employed explicitly considering all permutations. This limits the applicability of space based EDCG to smaller molecules or higher CG level (<9 CG sites). In the following a new algorithm is presented that can do such comparisons more efficiently and allows the extension of space based EDCG to larger proteins.

DEFINITIONS AND ALGORITHM DESIGN

To find a way to efficiently construct a permutation of CG site numbering that maximizes the overlap between different CG assignments, let us consider the following:

Definition 1. *Let $\mathcal{X} = \{x_i | i = 1, \ldots, N\}$ be the set of points that are to be clustered into different domains.*

Definition 2. *Let $\mathcal{M}^\alpha = \{M_i^\alpha \subset \mathcal{X} | i = 1, \ldots, n\}$ be a decomposition of $\mathcal{X}$ for all $\alpha = 1, \ldots, N_S$ such that*

$$\bigcup_{i=1}^{n} M_i^\alpha = \mathcal{X} \tag{C.1}$$

$$M_i^\alpha \cap M_j^\alpha = \emptyset . \tag{C.2}$$

Usually $\mathcal{X}$ is the set of C_α indices in the protein and the M_i^α are the sets of C_α's that are grouped into the CG site i in the α-th simulated annealing run of the space based EDCG.

Definition 3. *The function* $f_i^\alpha : \mathcal{X} \mapsto \{0,1\}$ *with*

$$f_i^\alpha = \begin{cases} 1 & x \in M_i^\alpha \\ 0 & x \notin M_i^\alpha \end{cases}$$

is called the member function of M_i^α.

Definition 4. *The overlap between two sets is*

$$\Omega_{ij}^{\alpha\alpha'} = \sum_{x\in\mathcal{X}} f_i^\alpha(x)\, f_j^{\alpha'}(x),\ i,j = 1,\dots,n\,.$$

This defines an overlap matrix $\Omega^{\alpha\alpha'} \in \mathbb{N}_0^{n\times n}$ *between the decompositions* $\mathcal{M}^\alpha$ *and* $\mathcal{M}^{\alpha'}$.

We are now interested in finding a permutation κ^* s.t.

$$\sum_{i=1}^{n} \Omega_{i\kappa^*(i)}^{\alpha\alpha'} \geqslant \sum_{i=1}^{n} \Omega_{i\kappa(i)}^{\alpha\alpha'}\,. \tag{C.3}$$

For all permutation $\kappa : \{1,\dots,n\} \mapsto \{1,\dots,n\}$. This permutation can be found in the following way:

Definition 5. *Let* $F^{\alpha\alpha'} : \{1,\dots,n\} \mapsto \{1,\dots,n\} \times \{1,\dots,n\}$ *be such that*

$$\Omega_{i\,F_{1,i}^{\alpha\alpha'}}^{\alpha\alpha'} \geqslant \Omega_{i\,F_{2,i}^{\alpha\alpha'}}^{\alpha\alpha'} \geqslant \cdots \geqslant \Omega_{i\,F_{n,i}^{\alpha\alpha'}}^{\alpha\alpha'}\,.$$

Then it follows that

$$\sum_{i=1}^{n} \Omega_{i\,F_{1,i}^{\alpha\alpha'}}^{\alpha\alpha'} \geqslant \sum_{i=1}^{n} \Omega_{i\kappa(i)}^{\alpha\alpha'},\ \forall\kappa\,. \tag{C.4}$$

However $F_{1,i}^{\alpha\alpha'}$ is not necessarily a permutation as it is not guaranteed that $F_{1,i}^{\alpha\alpha'} \neq F_{1,j}^{\alpha\alpha'}$ if $i \neq j$.

Because the starting point is the supremum of the overlap it is not possible that a larger overlap is obtained by moving a doubly occurring site to another already occupied site. Maximizing the overlap thus becomes equivalent to resolving the doubly occurring sites in a way that minimizes the loss to the overlap.

This means the double assignment of an index should be resolved in a way that results in the smallest loss to the overall overlap of the sequences. This is done by looking at the places $I_1,\dots,I_M$ for which $F_{1,I_k}^{\alpha\alpha'} = F_{1,I_l}^{\alpha\alpha'}$, $k,l = 1,\dots,M$ and picking I_{l^*} such that

$$F_{2,I_{l^*}}^{\alpha\alpha'} \geqslant F_{2,I_l}^{\alpha\alpha'},\ \forall l = 1,\dots,M\,.$$

Then the elements of $F_{\cdot,I_{l^*}}^{\alpha\alpha'}$ are shifted to the left (with periodic boundaries).

This can be repeated successively until no double assignments remain.

COMPUTATIONAL ASPECTS

The first computationally expensive part in this analysis is setting up of $\Omega^{\alpha\alpha'}$ which needs n^2 evaluations of the full sequence overlap. This could possibly be improved further, e.g. by setting the overlap to zero if the number of elements in M_i^{α} and $M_j^{\alpha'}$ is too different. If one could show that $\Omega^{\alpha\alpha'}$ can be transformed to diagonal form it would also help. This however needs a special proof because $\Omega^{\alpha\alpha'}$ is in general not symmetric. The improvement of scaling in this section is important because the evaluation of the full protein sequence will contribute significantly to the computational costs as it can be assumed that a large number of CG sites goes hand in hand with large proteins/complexes. Taking a fixed number of residues per CG site as an estimate for the operations needed to calculate each overlap shows that this operation contributes linearly to the scaling. With this assumption the first part of the program scales as n^3.

On top of the matrix construction the algorithm needs n sorting runs of an n member list. At the moment this is done using a simple bubble sort algorithm, giving an n^2 scaling. This scaling is good enough for present CG models with $< \mathcal{O}(100)$ CG sites. If larger numbers of CG sites are necessary better sorting algorithms can be used. Depending on the sorting algorithm the scaling for this part is in the range n^3 to $n^2 \log(n)$.

Next the scaling for the permutation search, should in the worst case be $n!$. But this is likely excluded by the fact that the sets M_i^{α} are disjoint and should combined be $\mathcal{X}$. The worst case would be that all indices are mapped to the same site in the first iteration. This in turn would mean that the overlap between all initial sets is largest with one final set, which should then contain the majority of all points. Admittedly this is not impossible but such an uneven distribution of residues among the CG sites should not occur for the minimized structures. The best case scaling would again be n^2 to make the check that all final indices are unique.

Overall the scaling of the whole program is likely on the order of n^3 for most use cases.

Under similar conditions the exhaustive search needs to try all possible $n!$ permutations and evaluate the whole protein sequence in each case. The overall scaling would accordingly be $n \cdot n!$. In the limit of small CG sites the exhaustive search might scale better, but the runtime in this case is dominated by overhead from setting up variables and data structures.

D

DETAILED DERIVATION OF THE CHARGE FITTING EQUATIONS AND RESULTING CHARGES

In the following the charge fitting equations briefly derived in chapter 7.3.2 are derived in more detail for various cases.

D.1 VACUUM & SCREENED ELECTROSTATICS

The easiest potential form is the case of vacuum electrostatics. In this case the effect of ions is ignored when the CG charges are determined and explicit counter charges need to be included in the CG simulation.

To determine the charges of the CG model the electrostatic vacuum potential is matched. For an isolated conformation of the atomistic model this is done by minimizing

$$\chi(\{\mathbf{r}_i\},\{\mathbf{R}_I\},\{q_i\},\{Q_I\}) = \int d^3r(\Phi^{AA}(\mathbf{r},t) - \Phi^{CG}(\mathbf{r},t))^2\,, \quad \text{(D.1)}$$

with respect to the CG charges Q_I. The electrostatic potentials are given by:

$$\Phi^{AA}(\mathbf{r},t) = \sum_{i=1}^{N_A} \frac{q_i}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{r}_i(t)|} \quad \text{(D.2)}$$

$$\Phi^{CG}(\mathbf{r},t) = \sum_{I=1}^{N_{CG}} \frac{Q_I}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{R}_I(t)|}\,. \quad \text{(D.3)}$$

Where N_A is the number of atoms with positions $r_i(t)$ and charges q_i in the all-atom simulations while N_{CG} is the number of CG sites with positions $R_I(t)$ and charges Q_I in the CG model. Since the all-atom and CG conformations as well as the q_i are fixed for a single conformation they are presently not written in the argument of χ.

To calculate χ in eq. (D.1) the function

$$f(\mathbf{x},\mathbf{y}) = \int d^3r \frac{1}{|\mathbf{r}-\mathbf{x}|}\frac{1}{|\mathbf{r}-\mathbf{y}|} \quad \text{(D.4)}$$

is considered first. The integral can be calculated by letting $\mathbf{x}$ be the origin and using spherical coordinates:

$$
\begin{aligned}
f(\mathbf{x},\mathbf{y}) &= \lim_{R\to\infty}\int_0^{2\pi} d\varphi \int_0^R dr \int_{-1}^1 d\cos\vartheta && \text{(D.5a)}\\
&\quad \frac{r^2}{r\sqrt{r^2+a^2-2\,a\,r\,\cos\vartheta}} && \text{(D.5b)}\\
&= 2\pi \lim_{R\to\infty}\int_0^R dr \int_{-1}^1 d\tau \frac{r}{\sqrt{r^2+a^2-2ar\tau}} && \text{(D.5c)}\\
&= \frac{2\pi}{a}\lim_{R\to\infty}\int_0^R dr\,(|r+a|-|r-a|) && \text{(D.5d)}\\
&= \frac{2\pi}{a}\lim_{R\to\infty}\left(2aR-a^2\right) && \text{(D.5e)}
\end{aligned}
$$

where $a = |\mathbf{x}-\mathbf{y}|$. The final result for $f(\mathbf{x},\mathbf{y})$ is thus

$$
f(\mathbf{x},\mathbf{y}) = \lim_{R\to\infty}\left[4\pi R - 2\pi|\mathbf{x}-\mathbf{y}|\right]. \tag{D.6}
$$

From this one can see that $f(\mathbf{x},\mathbf{y})$ has a divergent term, independent of $\mathbf{x}$ or $\mathbf{y}$.

With eq. (D.6) one can easily calculate the individual terms in eq. (D.1):

$$
\begin{aligned}
\int d^3r\,\Phi^{AA}(\mathbf{r},t)\Phi^{AA}(\mathbf{r},t) &= \sum_{i,j=1}^{N_A}\frac{q_i\,q_j}{(4\pi\varepsilon_0)^2}\,f(\mathbf{r_i}(t),\mathbf{r_j}(t))\\
\int d^3r\,\Phi^{CG}(\mathbf{r},t)\Phi^{CG}(\mathbf{r},t) &= \sum_{L,M=1}^{N_{CG}}\frac{Q_L\,Q_M}{(4\pi\varepsilon_0)^2}\,f(\mathbf{R_i}(t),\mathbf{R_j}(t))\\
\int d^3r\,\Phi^{AA}(\mathbf{r},t)\Phi^{CG}(\mathbf{r},t) &= \sum_{i=1}^{N_A}\sum_{I=1}^{N_{CG}}\frac{q_i\,Q_I}{(4\pi\varepsilon_0)^2}\,f(\mathbf{r_i}(t),\mathbf{R_I}(t)).
\end{aligned}
$$

And thus

$$
\begin{aligned}
\chi(\{Q_I\}) = (4\pi\varepsilon_0)^2 \lim_{R\to\infty}\Bigg[& \\
& 4\pi R\left(\sum_{i,j=1}^{N_A} q_i\,q_j + \sum_{L,M=1}^{N_{CG}} Q_L\,Q_M - 2\sum_{i=1}^{N_A}\sum_{I=1}^{N_{CG}} q_i\,Q_I\right)\\
& -2\pi\sum_{i,j=1}^{N_A} q_i\,q_j\left|\mathbf{r}_i(t)-\mathbf{r}_j(t)\right|\\
& -2\pi\sum_{L,M=1}^{N_{CG}} Q_L\,Q_M\,|\mathbf{R}_L(t)-\mathbf{R}_M(t)|\\
& +4\pi\sum_{i=1}^{N_A}\sum_{I=1}^{N_{CG}} q_i\,Q_I\,|\mathbf{R}_I(t)-\mathbf{r}_i(t)|\Bigg].
\end{aligned} \tag{D.7}
$$

For $\mathcal{X}$ to be a useful measure of similarity in the electrostatic potentials the prefactor of the diverging term in eq. (D.7) has to be equal to zero. The prefactor can be written as:

$$\sum_{i,j=1}^{N_A} q_i\, q_j + \sum_{L,M=1}^{N_{CG}} Q_L\, Q_M - 2\sum_{i=1}^{N_A}\sum_{I=1}^{N_{CG}} q_i\, Q_I = \left(\sum_{i=1}^{N_A} q_i - \sum_{I=1}^{N_{CG}} Q_I\right)^2 \tag{D.8}$$

so that $\mathcal{X}$ is well behaved if the total charge of the CG model is the same as the total charge of the atomistic model. Considering the long range nature of the vacuum electrostatic interaction it is not surprising that the monopole moments have to be equal. After all the long range behavior is dominated by the monopole moment of the charge distribution. This conclusion could have been derived without further considerations from Gauss' law.

To obtain CG charges that are independent of the chosen configuration the thermal average of eq. (D.1) with respect to the all atom trajectory is considered:

$$\chi^2_\Phi(\{Q_I\}) = \langle \mathcal{X}(\{\mathbf{r}_i\},\{M_I(\{\mathbf{r}_i\})\},\{Q_I\})\rangle_{AA}\ . \tag{D.9}$$

Where the operators $\{M_I\}$ are the mappings from an all-atom configuration to a CG configuration and the all-atom thermal average is performed according to

$$\begin{aligned}\langle f(\{\mathbf{r}_i\})\rangle_{AA} &= \frac{\int d^3r_1 \ldots d^3r_{N_A}\, f(\{\mathbf{r}_i\})\, \exp\left[-\frac{U(\{\mathbf{r}_i\})}{k_B T}\right]}{\int d^3r_1 \ldots d^3r_{N_A}\, \exp\left[-\frac{U(\{\mathbf{r}_i\})}{k_B T}\right]} \\ &= \frac{1}{Z}\int d^{3N_A}r\, f(\underline{r})\, \exp\left[-\frac{U(\underline{r})}{k_B T}\right]\ . \end{aligned} \tag{D.10}$$

Using equations (D.7), (D.8), and (D.10) equation (D.9) can be written as

$$\begin{aligned}
\chi^2_\Phi(\{Q_I\}) &= -\frac{2\pi}{Z}\int d^{3N_A} r\, e^{-\frac{U(\underline{r})}{k_B T}} \left[\sum_{i,j=1}^{N_A} \frac{q_i\, q_j}{(4\pi\varepsilon_0)^2} \left|\mathbf{r}_i - \mathbf{r}_j\right| \right. \\
&\quad + \sum_{L,M=1}^{N_{CG}} \frac{Q_L\, Q_M}{(4\pi\varepsilon_0)^2} \left|M_L(\underline{r}) - M_M(\underline{r})\right| \\
&\quad \left. -2 \sum_{I=1}^{N_{CG}} \sum_{i=1}^{N_A} \frac{Q_I\, q_i}{(4\pi\varepsilon_0)^2} \left|M_I(\underline{r}) - \mathbf{r}_i)\right| \right] \\
&= -\sum_{i,j=1}^{N_A} \frac{q_i\, q_j}{(4\pi\varepsilon_0)^2} \langle 2\pi \left|\mathbf{r}_i - \mathbf{r}_j\right| \rangle_{AA} \\
&\quad +2 \sum_{I=1}^{N_{CG}} \sum_{i=1}^{N_A} \frac{Q_I\, q_i}{(4\pi\varepsilon_0)^2} \langle 2\pi \left|M_I(\underline{r}) - \mathbf{r}_i)\right| \rangle_{AA} \qquad \text{(D.11)} \\
&\quad - \sum_{L,M=1}^{N_{CG}} \frac{Q_L\, Q_M}{(4\pi\varepsilon_0)^2} \langle 2\pi \left|M_L(\underline{r}) - M_M(\underline{r})\right| \rangle_{AA} .
\end{aligned}$$

The minimum is determined by solving the system of equations

$$\frac{\partial \chi^2_\Phi(\{Q_I\})}{\partial Q_\alpha} = 0,\ \forall \alpha\,, \qquad \text{(D.12)}$$

which is equivalent to solving

$$\sum_{I=1}^{N_{CG}} \langle \left|M_\alpha(\underline{r}) - M_I(\underline{r})\right| \rangle_{AA}\, Q_I = \sum_{i=1}^{N_A} \langle \left|M_\alpha(\underline{r}) - \mathbf{r}_i\right| \rangle_{AA}\, q_i \qquad \text{(D.13)}$$

for all α. Equation (D.13) can be written in matrix form:

$$\begin{aligned}
\underline{\underline{A}}\,\underline{Q} &= \underline{\underline{B}}\,\underline{q} && \text{(D.14a)} \\
A_{ij} &= \langle \left|M_i(\underline{r}) - M_j(\underline{r})\right| \rangle_{AA} = A_{ji} && \text{(D.14b)} \\
&\quad \underline{\underline{A}} \in \mathbb{R}^{N_{CG} \times N_{CG}} \\
B_{ij} &= \langle \left|M_\alpha(\underline{r}) - \mathbf{r}_i\right| \rangle_{AA} && \text{(D.14c)} \\
&\quad \underline{\underline{B}} \in \mathbb{R}^{N_{CG} \times N_A}\,.
\end{aligned}$$

Assuming ergodicity of the system the esemble average can be written as a time average:

$$\begin{aligned}
\sum_{I=1}^{N_{CG}} \int_0^{t_{max}} \frac{dt \left|M_\alpha(\underline{r}(t)) - M_I(\underline{r}(t))\right|}{t_{max}}\, Q_I &= \\
\sum_{i=1}^{N_A} \int_0^{t_{max}} \frac{dt \left|M_\alpha(\underline{r}(t)) - \mathbf{r}_i(t)\right|}{t_{max}}\, q_i \qquad &. \qquad \text{(D.15)}
\end{aligned}$$

This time average can be evaluted using MD simulations in which case the time integration turns into a summation over the N_t collected simulation frames:

$$\sum_{I=1}^{N_{CG}} \sum_{n=1}^{N_t} |M_\alpha(\underline{r}(t_n)) - M_I(\underline{r}(t_n))| \; Q_I \quad = \quad \sum_{i=1}^{N_A} \sum_{n=1}^{N_t} |M_\alpha(\underline{r}(t_n)) - \mathbf{r}_i(t_n)| \; q_i \qquad , \tag{D.16}$$

or in matrix form:

$$\underline{\underline{A}}' \, \underline{Q} \;=\; \underline{\underline{B}}' \, \underline{q} \tag{D.17a}$$

$$A'_{ij} \;=\; \sum_{n=1}^{N_t} \left| M_i(\underline{r}(t_n)) - M_j(\underline{r}(t_n)) \right| = A'_{ji} \tag{D.17b}$$

$$B'_{ij} \;=\; \sum_{n=1}^{N_t} \left| M_i(\underline{r}(t_n)) - \mathbf{r}_j(t_n) \right| . \tag{D.17c}$$

In the following an approach is taken that is also used in force matching [205]. Equation (D.16) is fulfilled if the equality holds for every term in the sum:

$$\sum_{I=1}^{N_{CG}} |M_\alpha(\underline{r}(t_n)) - M_I(\underline{r}(t_n))| \; Q_I \quad = \quad \sum_{i=1}^{N_A} |M_\alpha(\underline{r}(t_n)) - \mathbf{r}_i(t_n)| \; q_i, \qquad \forall n, \alpha \,, \tag{D.18}$$

or in matrix form:

$$\underline{\underline{\widetilde{A}}} \, \underline{Q} \;=\; \underline{\underline{\widetilde{B}}} \, \underline{q} \tag{D.19a}$$

$$\widetilde{A}_{ij} \;=\; \left| M_i(\underline{r}(t_n)) - M_j(\underline{r}(t_n)) \right| \tag{D.19b}$$

$$\underline{\underline{\widetilde{A}}} \in \mathbb{R}^{(N_t \cdot N_{CG}) \times N_{CG}}$$

$$\widetilde{B}_{ij} \;=\; \left| M_i(\underline{r}(t_n)) - \mathbf{r}_j(t_n) \right| \tag{D.19c}$$

$$\underline{\underline{\widetilde{B}}} \in \mathbb{R}^{(N_t \cdot N_{CG}) \times N_A} .$$

The quantities $\underline{\underline{\widetilde{A}}}$ and $\underline{\underline{\widetilde{B}}}$ are vectors of matrices.
This is equivalent to $N_{CG} \cdot N_t$ equations which overdetermine the N_{CG} Q_I if more than one time frame is considered. Thus eq. (D.18) has to be solved in a least squares way. This implies that in general the solution will not be exact, so that the resulting $\{Q_I\}$ are not the optimal charges for eq. (D.15).

The correction to the optimal charges can be calculated as follows: Multiplying eq. (D.19a) from the left with $\underline{\underline{\widetilde{A}}}^{T}$ and eq. (D.17a) with $\underline{\underline{A'}}^{T} = \underline{\underline{A'}}$ and comparing the resulting terms one finds

$$\begin{aligned}
\underline{\underline{A'}}\,\underline{\underline{A'}} &= \underline{\underline{\widetilde{A}}}^{T}\underline{\underline{\widetilde{A}}} + \underline{\underline{\Xi}}_1 && \text{(D.20a)}\\
\Xi_{1\,ij} &= \sum_{L=1}^{N_{CG}} \sum_{n_1 \neq n_2} |M_i(\underline{r}(t_{n_1})) - M_L(\underline{r}(t_{n_1}))| \\
&\quad \times \left|M_L(\underline{r}(t_{n_2})) - M_j(\underline{r}(t_{n_2}))\right| \\
\underline{\underline{A'}}\,\underline{\underline{B'}} &= \underline{\underline{\widetilde{A}}}^{T}\underline{\underline{\widetilde{B}}} + \underline{\underline{\Xi}}_2 && \text{(D.20b)}\\
\Xi_{2\,ij} &= \sum_{L=1}^{N_{CG}} \sum_{n_1 \neq n_2} |M_i(\underline{r}(t_{n_1})) - M_L(\underline{r}(t_{n_1}))| \\
&\quad \times \left|M_L(\underline{r}(t_{n_2})) - \mathbf{r}_j(t_{n_2})\right| .
\end{aligned}$$

The CG charges are split into the solution $\underline{Q}_0$ of (D.19a) and a correction $\underline{\delta Q}$. Using the equations (D.20a) and (D.20b), equation (D.17a) can be written as:

$$\underline{\underline{\widetilde{A}}}^{T} \underbrace{\left(\underline{\underline{\widetilde{A}}}\,\underline{Q}_0 - \underline{\underline{\widetilde{B}}}\,\underline{q}\right)}_{=\underline{\rho}(t_n)} + \underbrace{\left(\underline{\underline{\Xi}}_1\,\underline{Q}_0 - \underline{\underline{\Xi}}_2\,\underline{q}\right)}_{=\underline{c}} + \underline{\underline{A'}}\,\underline{\underline{A'}}\,\underline{\delta Q} = 0 . \quad \text{(D.21)}$$

The two quantities $\underline{\rho}(t_n)$ and $\underline{c}$ are related and would be zero if the overdetermined system in eq. (D.19a) had an exact solution. In this case it would follow that $\underline{\delta Q}$ is also zero (assuming that $\underline{\underline{A'}}$ can be inverted, so that eq. (D.17a) has a unique solution). In general this will not be the case and $\underline{\rho}(t_n)$ and $\underline{c}$ are given by

$$\begin{aligned}
\rho_\alpha(t_n) &= \sum_{I=1}^{N_{CG}} |M_\alpha(\underline{r}(t_n)) - M_I(\underline{r}(t_n))|\, Q_{0\,I} \\
&\quad - \sum_{i=1}^{N_A} |M_\alpha(\underline{r}(t_n)) - \mathbf{r}_i(t_n)|\, q_i && \text{(D.22)}\\
c_\alpha &= \sum_{L=1}^{N_{CG}} \sum_{n_1 \neq n_2} |M_\alpha(\underline{r}(t_{n_1})) - M_L(\underline{r}(t_{n_1}))|\, \rho_L(t_{n_2}) .
\end{aligned}$$

From this follows

$$\underline{\underline{\widetilde{A}}}^{T}\underline{\rho}(t_n) + \underline{c} = \underline{\underline{A'}}\left(\sum_{n=1}^{N_t} \underline{\rho}(t_n)\right) \quad \text{(D.23)}$$

so that the correction to the charges are given by:

$$\underline{\delta Q} = \underline{\underline{A'}}^{-1}\left(\sum_{n=1}^{N_t} \underline{\rho}(t_n)\right) . \quad \text{(D.24)}$$

This means finding a mean square solution to eq. (D.19a) can produce the correct CG charges according to eq. (D.17a) even if the solution is not exact in every frame. Inserting the definition in

eq. (D.22) into the condition $\sum_n \underline{\rho}(t_n) = \underline{0}$ reproduces eq. (D.16).

To enforce the equality of the total charges during the calculation a Langrane multiplier approach is taken:

$$L(\underline{Q}, \lambda) = \chi^2_\Phi(\underline{Q}) + \lambda \left(\sum_{J=1}^{N_{CG}} Q_I - \sum_{i=1}^{N_A} q_i \right) , \tag{D.25}$$

so that the condition for a minimum are

$$\left(\frac{\partial}{\partial \underline{Q}}, \frac{\partial}{\partial \lambda} \right)^T L(\underline{Q}, \lambda) = \underline{0} \tag{D.26a}$$

$$\Updownarrow$$

$$8\varepsilon_0 \lambda = -\widetilde{\lambda} = -\left(\underline{\underline{A}}\,\underline{Q}\right)_\alpha + \left(\underline{\underline{B}}\,\underline{q}\right)_\alpha , \forall \alpha \tag{D.26b}$$

$$\sum_{I=1}^{N_{CG}} Q_I = \sum_{i=1}^{N_A} q_i . \tag{D.26c}$$

This can be solved by considering:

$$\sum_{L=1}^{N_{CG}} Q_L = \sum_{L=1}^{N_{CG}} \left(\underline{\underline{A}}^{-1} \underline{\underline{B}}\,\underline{q}\right)_L + \widetilde{\lambda} \sum_{L,M=1}^{N_{CG}} \left(\underline{\underline{A}}^{-1}\right)_{LM} = \sum_{i=1}^{N_A} q_i \tag{D.27}$$

This allows the determination of $\widetilde{\lambda}$ as

$$\widetilde{\lambda} = \frac{\sum_{i=1}^{N_A} q_i - \sum_{L=1}^{N_{CG}} \left(\underline{\underline{A}}^{-1} \underline{\underline{B}}\,\underline{q}\right)_L}{\sum_{L,M=1}^{N_{CG}} \left(\underline{\underline{A}}^{-1}\right)_{LM}} . \tag{D.28}$$

Using the shorthand

$$\Omega_I = \sum_{L=1}^{N_{CG}} \left(\underline{\underline{A}}^{-1}\right)_{IL} \quad , \quad \Gamma_I = \sum_{L=1}^{N_{CG}} \sum_{i=1}^{N_A} \left(\underline{\underline{A}}^{-1}\right)_{IL} B_{Li}\, q_i \tag{D.29}$$

the CG charges are then given by

$$Q_I = \Gamma_I + \Omega_I \frac{\sum_{i=1}^{N_A} q_i - \sum_{L=1}^{N_{CG}} \Gamma_L}{\sum_{L=1}^{N_{CG}} \Omega_L} . \tag{D.30}$$

Summerizing the result one has:

$$\chi^2_\Phi = \frac{2\pi}{4\varepsilon_0}\left(2\,\underline{Q}^T\underline{\underline{B}}\,\underline{q} - \underline{Q}^T\underline{\underline{A}}\,\underline{Q} - \underline{q}^T\underline{\underline{C}}\,\underline{q}\right) \tag{D.31a}$$

$$A_{ij} = \frac{1}{N_t}\sum_{n=1}^{N_t}\left|M_i(\underline{r}(t_n)) - M_j(\underline{r}(t_n))\right| \in \mathbb{R}^{N_{CG}\times N_{CG}} \tag{D.31b}$$

$$B_{ij} = \frac{1}{N_t}\sum_{n=1}^{N_t}\left|M_i(\underline{r}(t_n)) - \mathbf{r}_j(t_n)\right| \in \mathbb{R}^{N_{CG}\times N_A} \tag{D.31c}$$

$$C_{ij} = \frac{1}{N_t}\sum_{n=1}^{N_t}\left|\mathbf{r}_i(t_n) - \mathbf{r}_j(t_n)\right| \in \mathbb{R}^{N_A\times N_A} \tag{D.31d}$$

$$Q_i = \Gamma_i + \Omega_i\frac{\sum_{l=1}^{N_A} q_l - \sum_{L=1}^{N_{CG}}\Gamma_L}{\sum_{L=1}^{N_{CG}}\Omega_L} \tag{D.31e}$$

$$\Omega_i = \sum_{l=1}^{N_{CG}}\left(\underline{\underline{A}}^{-1}\right)_{il} \tag{D.31f}$$

$$\Gamma_i = \sum_{l=1}^{N_{CG}}\sum_{k=1}^{N_A}\left(\underline{\underline{A}}^{-1}\right)_{il} B_{lk}\,q_k \tag{D.31g}$$

As it might be advantageous to perform the CG simulations in a continuum solvent model having explicit counter charges would be counter productive. Therefore it is more suitable to assume a screened electrostatic potential. The calculations for this case proceed along the same lines and are outlined in section 7.3.2 of the main text. The result of this calculation are the operators defined in eq. (7.16b). In the case of weak screening ($\lambda \to \infty$) the results of the screened and the vacuum case are equivalent. This can be seen by expanding eq. (7.16b) in λ^{-1}:

$$A_{ij} = \lambda - \left\langle\left|\mathbf{r}_i(t_n) - \mathbf{R}_j(t_n)\right|\right\rangle + \mathcal{O}(\lambda^{-1})\,. \tag{D.32}$$

The first term is the long range divergence that cancels in the residual if the overall charge is conserved, the second term is just the result from the vacuum calculations and all higher order terms are neglected.

D.1.1 *Periodic Boundary Conditions*

The treatment presented above is sufficient to match the electrostatic potential created by a protein. However to treat surfaces the periodic boundary conditions have to be respected in order to avoid finite size effects at the borders of the simulation cell.

Since the periodic boundary conditions lead to a periodic charge distribution the resulting potential can be calculated by solving the screened Poisson equation

$$\left(-\Delta + \frac{1}{\lambda^2}\right)\phi(\mathbf{r}) = 4\pi\varepsilon_0\rho(\mathbf{r}) = 4\pi\varepsilon_0\sum_i q_i\delta(\mathbf{r}-\mathbf{r}_i) \quad \text{(D.33)}$$

using a Fourier series representation of ρ and ϕ:

$$\sum_{\mathbf{n}\in\mathbb{Z}^3}\left(k_{\mathbf{n}}^2 + \frac{1}{\lambda^2}\right)\varphi_{\mathbf{n}}e^{ik_{\mathbf{n}}\cdot\mathbf{r}} = 4\pi\varepsilon_0\sum_i\sum_{\mathbf{n}\in\mathbb{Z}^3} q_i e^{ik_{\mathbf{n}}\cdot\mathbf{r}}\,. \quad \text{(D.34)}$$

The electrostatic potential produced by the periodic charge distribution is therefore

$$\phi(\mathbf{r}) = \sum_i\sum_{\mathbf{n}\in\mathbb{Z}^3}\frac{q_i}{4\pi\varepsilon_0}\frac{e^{ik_{\mathbf{n}}\cdot\mathbf{r}}}{k_{\mathbf{n}}^2 + \frac{1}{\lambda^2}}\,. \quad \text{(D.35)}$$

This makes it easy to evalute the terms of the form $\int d^3r|\phi(\mathbf{r})|^2$ using the Fourier representation of the δ-function:

$$\int d^3r|\phi(\mathbf{r})|^2 = \sum_{i,j}\frac{q_i\,q_j}{(4\pi\varepsilon_0)^2}\sum_{\mathbf{n}\in\mathbb{Z}^3}\frac{e^{ik_{\mathbf{n}}\cdot(\mathbf{r}_i-\mathbf{r}_i)}}{\left(k_{\mathbf{n}}^2 + \frac{1}{\lambda^2}\right)^2}\,. \quad \text{(D.36)}$$

The correct geometry of the box influences the result in this kind of calculation [351] and because typical simulation systems have finite cell dimensions and are typically not spherical the summation in eq. (D.36) can not be performed analytically. A numerical evaluation of the summation is in principal possible, but costly.

D.2 MIXING SCREENED AND VACUUM ELECTROSTATICS

D.2.1 *Derivation*

So far only fixed atomistic charges have been considered. In some cases the atomistic charges might change during the simulation. This is the case, e.g. for a CG model of the mica surface. The literature and all atom simulation of fibrinogen adsorption on mica suggest that charge effects play an important role for the interaction of fibrinogen with this surface. It seems prudent to include electrostatic interactions into the CG model of such a surface to capture these effect. Surfaces like mica have a counter ion layer at the water interface that can be replaced by mobile ions from solution. Due to defects in the mica structure, the ions might not be uniformly distributed along the surface. Including the surface ions with the bulk ions distorts the surface electrostatics while including the ions that replace them introduces a certain arbitrariness when deciding which ion is at the surface and which is in solution. This problem can be addressed by mixing an unscreened electrostatic potential for the atomistic

simulations with a screened electrostatic interaction for the CG model.

In the case of mixed vacuum and screened electrostatics the residual is given by

$$\chi^2_\Phi = -\frac{2\pi}{(4\pi\varepsilon_0\varepsilon_{H_2O})^2}\left(\underline{q}\,\underline{\underline{C}}\,\underline{q} + \underline{Q}\,\underline{\underline{A}}\,\underline{Q} - 2\underline{Q}\,\underline{\underline{B}}\,\underline{q}\right), \tag{D.37}$$

with

$$A_{ij} = -\frac{\lambda}{\pi N_t}\sum_{n=1}^{N_t}\exp\left[-\frac{\left|\mathbf{r}_i(t_n)-\mathbf{R}_j(t_n)\right|}{\lambda}\right] \tag{D.38a}$$

$$B_{ij} = -\frac{2\lambda^2}{N_t}\sum_{n=0}^{N_t}\frac{1-\exp\left[-\frac{\left|\mathbf{r}_i(t_n)-\mathbf{R}_j(t_n)\right|}{\lambda}\right]}{\left|\mathbf{r}_i(t_n)-\mathbf{R}_j(t_n)\right|} \tag{D.38b}$$

$$C_{ij} = \frac{1}{N_t}\sum_{n=1}^{N_t}\left|\mathbf{r}_i(t_n)-\mathbf{r}_j(t_n)\right|. \tag{D.38c}$$

All of these operators give the vacuum result in the limit $\lambda \to \infty$. It is important to note however that for finite λ the divergence term from the unscreened atomistic potential is not canceled by the coarse grained contribution, which is now short ranged. The resulting term

$$\lim_{R\to\infty}\sum_{i,j} q_i q_j R = \lim_{R\to\infty}\left(\sum_i q_i\right)^2 R$$

however cancels if the system does not posses a net charge.

This approach significantly increases finite size effects. If, e.g. the two sides of a unit cell are not neutral neglecting the periodic boundary conditions will give rise to a polarization across the box. As this is a long range effect, the short ranged coarse grained potential can only reproduce it by a charge gradient across the cell. This artificial gradient obscures the correct CG charges. To avoid this, a periodic treatment of the atomistic potential is necessary.

One suggestions might be to define the box in such a way that the borders are neutral. In cases like mica the surface structure is too complicated to allow such a systematic cut. Even if this was possible, such an approach would give rise to complicated boundary conditions.

Again a Fourier representation leads to operators of the form of eq. (D.36), with the exception that the denominator in the $\mathbf{k_n}$ sum is not squared. Unfortunately this does not make the summation easier. However, it allows for a semi-numerical approach. In this approach the atomistic potential is evaluated numerically using available software, e.g. DelPhi [352, 353]. This numerical evaluation needs to treat the periodic boundary conditions correctly. The CG electric field is short ranged, so that a cutoff can be used. This means the system analyzed has to be big enough

so that the region of interest is far enough from the boundary (far enough being a function of λ). If this is not the case a full numerical iterative procedure has to be used. In such a procedure the CG charges would be iteratively refined to match the tabulated CG electric field trajectory to the atomistic electric field trajectory. This requires a recalculation of the CG electric field trajectory in each iteration and is therefore costly.

As the focus is to fit surface charges, it is always possible to simulate a big enough slab and then focus on the inner region. Therefore the analysis can be limited to the case where only the atomistic electric field is tabulated, s.t.

$$\Phi^{AA}(\mathbf{r},t) = \phi(\mathbf{r}_i) = \phi_i, r \in V_i \tag{D.39}$$

$$\Phi^{CG}(\mathbf{r},t) = \sum_{I=1}^{N_{CG}} \frac{Q_I e^{|\mathbf{r}-\mathbf{R}_I(t)|/\lambda}}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{R}_I(t)|}. \tag{D.40}$$

The resulting residual is

$$\begin{aligned}\chi^2_\Phi = {} & V\sum_i \langle\phi_i^2\rangle + \frac{2\pi\lambda}{(4\pi\varepsilon_0)^2}\sum_{L,M} Q_L \left\langle e^{-|\mathbf{R}_L-\mathbf{R}_M|/\lambda}\right\rangle Q_M \\ & -\frac{V}{2\pi\varepsilon_0}\sum_I\sum_i \left\langle \frac{\phi_i e^{-|\mathbf{R}_I-\mathbf{r}_i|/\lambda}}{|\mathbf{r}_i-\mathbf{R}_I|}\right\rangle Q_I,\end{aligned} \tag{D.41}$$

where $\mathbf{r}_i$ is the center of V_i. In line with the fully numerical approach it was assumed that the CG potential can be seen as constant in each cube of volume V associated with the numerics of the atomistic potential.

The set of CG charges minimizing eq. (D.41) is determined by the solution of

$$\sum_L \left\langle e^{-|\mathbf{R}_L-\mathbf{R}_\alpha|/\lambda}\right\rangle Q_L = 2\frac{V\varepsilon_0}{\lambda}\left\langle \sum_i \phi_i \frac{e^{-|\mathbf{R}_\alpha-\mathbf{r}_i|/\lambda}}{|\mathbf{r}_i-\mathbf{R}_\alpha|}\right\rangle. \tag{D.42}$$

Due to the rigid nature of the surface the centers of mass of the CG beads barely move. It is therefore reasonable to take them as fixed in the CG model. In this case eq. (D.42) simplifies to

$$\sum_L e^{-|\mathbf{R}_L-\mathbf{R}^0_\alpha(0)|/\lambda} Q_L = 2\frac{V\varepsilon_0}{\lambda}\sum_i \langle\varphi_i\rangle \frac{e^{-|\mathbf{R}^0_\alpha-\mathbf{r}_i|/\lambda}}{|\mathbf{R}^0_\alpha-\mathbf{r}_i|}, \tag{D.43}$$

where $\mathbf{R}^0_\alpha$ is the initial ($t=0$) position of the CG bead α.

D.2.2 *Charge Structure of a Model Surface*

The atomistic simulations of the mica surface represent an infinite surface with the mica slab connecting across periodic boundaries. Without the periodic boundaries, uncompensated charges exist on either side of the box giving rise to an artificial dipole moment discussed above. Extending the system that is fitted to the periodic boundaries means that the assumption of a vanish-

ing field at the box boundaries can no longer be used and eq. (D.43) has to be used.

A $10.4nm \times 10.7nm \times 10.1nm$ box of a two layer mica sheet surrounded by physiological salt solution was used as benchmark system. The box was minimized for 5000 steps and equilibrated for 1ns. For the calculation of the average field NPT MD trajectories of a length of $2 \times 10ns$ were collected. The CG level was chosen close to the resolution of the CG model for fibrinogen. The model closest to this resolution would have three beads in the super cell shown in Fig. 19. To preserve the shape better a model with four beads was chosen (315 atoms per bead instead of ≈422 atoms per bead in the fibrinogen model). Monitoring the projected movement of these beads along the simulated fibrogen adsorption trajectories validates the assumption that the beads are essentially stationary throughout the simulation.

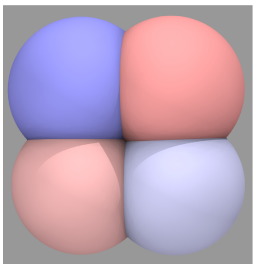

Figure 154: A CG model of the mica surface. The CG beads are colored according to their charge (with red being the most negative (-0.1e) and blue the most positive values (0.6e)). A screened electrostatic interaction is assumed for the atomistic and coarse grained interactions.

The calculation of $\{\langle\phi_i\rangle\}_i$ was attempted with APBS and DelPhi, neither of which was successful. Since ions are included explicitly in these simulations the ion concentration in the Poisson-Boltzmann equation was set to zero. The explicit solvent was removed and replaced by a dielectric medium with a relative permittivity of 80, representing water. The APBS software does not support a full treatment of periodic boundary conditions and is thus not helpful. DelPhi does in principle support a full treatment of the periodic boundary conditions. However, if the grid was chosen such that the mica surface extends to the border of the grid the solver diverged. Different parameters and convergence criteria were used to no avail. The identification of a better solver would be the obvious next step.

To get an impression of the kind of mica model that can be expected from charge fitting, a less realistic approach was taken. An isolated two sheet mica slab (originally constrcuted for the fibrinogen adsorption simulations) was taken as input for the charge fitting. For the CG as well as the atomistic model screened electrostatic interactions were assumed. A two super cell wide border region was not considered for the analysis to avoid boundary effects. This border region is 3.9nm wide and thus much larger than the screening length ($\lambda \approx 0.8nm$) in physiological solution. The resulting 2×7 super cells showed an overall very similar charge structure shown in Fig. 154. The exact charges vary across the box. The checkerboard pattern of the defect structure shown in Fig. 19 is clearly reproduced. The values also show that the mica surface displays a positively charged character as a consequence of the counter ion layer.

D.3 RESULTING CHARGES OF THE DIFFERENT APPROACHES

In the preceding section as well as 7.3.2 several methods for the determination of the charges have been proposed. Each of them leads to a different assignment of the CG charges. The resulting

charges for the different methods are given in Table 29 for the CG beads in one protomer. The second is symmetric.

Region	screened	screened +conserved	unscreened	summation
Fp	0.147e	−0.098e	−0.774e	4.00e
Disulfide knot	−0.152e	−0.380e	−0.557e	−3.00e
Coiled-Coil	−0.354e	−0.619e	−0.450e	−1.00e
Coiled-Coil	−0.670e	−0.936e	−1.459e	−5.00e
Coiled-Coil	0.147e	−0.091e	0.227e	3.00e
Hinge	0.317e	0.107e	0.709e	3.00e
Hinge	−0.436e	−0.654e	−0.910e	−2.00e
Coiled-Coil	−0.283e	−0.531e	−0.380e	−1.00e
Coiled-Coil	−0.168e	−0.432e	−0.142e	−2.00e
Coiled-Coil	0.200e	−0.072e	0.633e	2.00e
Coiled-Coil	0.016e	−0.255e	0.331e	2.00e
Coiled-Coil	−0.404e	−0.669e	−0.126e	−2.00e
Coiled-Coil	−0.091e	−0.324e	−0.092e	−4.00e
βC	−0.054e	−0.296e	−0.181e	1.84e
βC	−0.635e	−0.878e	−0.464e	0.00e
βC	−0.363e	−0.589e	−0.926e	1.00e
βC	−0.547e	−0.789e	−0.941e	−1.00e
γC	−0.829e	−1.067e	−2.843e	−4.00e
γC	0.288e	0.047e	1.570e	3.00e
γC	−0.736e	−0.978e	−2.497e	−2.00e
γC	−0.365e	−0.605e	−1.212e	−1.00e
γC	0.041e	−0.197e	0.470e	2.00e
Glycans	−0.457e	−0.747e	−1.374e	−1.84e

Table 29: Assignment of charges to the CG beads in a protomer with various methods

E

CALCULATIONS AND RESULTS FOR THE PHENOMENOLOGICAL MODEL

For the numerical implementation of the model it is necessary to introduce a certain scale in which the size of the molecule is measured. This scale is set in the following by measuring distances in units of l_2. Accordingly

$$\xi = \frac{l_1}{l_2} \tag{E.1}$$

is defined.

In the following the constraints for an adsorbed model with hard body interactions are derived and limiting cases are discused. Later approaches for the calculation of $P(\alpha)$ from $P(\gamma, \varphi)$ are discussed. Lastly numerically determined distributions $P_{ij}(\alpha)$ are presented together with the parameters resulting from a least squares fit to experiments.

E.1 DERIVATION OF CONSTRAINTS

The condition of an adsorbed state as well as the hard body interactions between the model and the surface impose the constraints

$$R_D < h_{D,i} < h_{max,D},\ i = 1,2 \tag{E.2a}$$
$$R_E < h_0 < h_{max,E} \tag{E.2b}$$
$$0 < h_0 - R_C \pm \xi \sin(\vartheta) \quad , \tag{E.2c}$$

where $h_{D,i}$ is the height of the D-regions above the surface. Equation (E.2c) ensures that the hinges do not intersect the surface.

To determine the distance of the globular regions from the surface one needs to know their coordinates. These are determined easily in solution, when the model is symmetric under rotation around its long axis. This symmetry makes cylindrical coordinates an appropriate choice for the calculation of the coordinates of the centers of the D-regions in relation to the E-region. The E-region is taken as the origin for these calculations with the long axis of the molecule pointing along the z-axis. Let $\widetilde{\mathbf{r}}_i$, $i = 1, 2$ be

the vectors from the E-region to the center of the D-regions. In the cylindrical coordinate system, these vectors are given by

$$\widetilde{\mathbf{r}}_i = \begin{pmatrix} \sin(\gamma_i)\,\cos(\varphi_i) \\ \sin(\gamma_i)\,\sin(\varphi_i) \\ \pm(\xi - \cos(\gamma_i)) \end{pmatrix}. \tag{E.3}$$

To transform to the system depicted in Fig. 92 first a rotation around the x-axis by an angle $\vartheta + \frac{\pi}{2}$ is performed before the molecule as a whole is shifted by h_0 in z-direction. The coordinates in the new system are

$$\mathbf{r}_i = R_x(\vartheta + \frac{\pi}{2})\widetilde{\mathbf{r}}_i = \begin{pmatrix} \sin(\gamma_i)\,\cos(\varphi_i) \\ \sin(\vartheta)\,\sin(\gamma_i)\,\cos(\varphi_i) \mp \cos(\vartheta)(\xi - \cos(\gamma_i)) \\ \cos(\vartheta)\,\sin(\gamma_i)\,\sin(\varphi_i) \pm \sin(\vartheta)(\xi - \cos(\gamma_i)) \end{pmatrix}. \tag{E.4}$$

Accordingly, the distances between the surface and the D-regions are

$$h_{D,1} = h_0 - \xi\sin(\vartheta) + (\mathbf{r}_1)_z \tag{E.5a}$$
$$h_{D,2} = h_0 + \xi\sin(\vartheta) + (\mathbf{r}_2)_z \tag{E.5b}$$

and the hinges are located at

$$\mathbf{h}_i = \begin{pmatrix} 0 \\ \pm\xi\,\cos(\vartheta) \\ h_0 \mp \xi\,\sin(\vartheta) \end{pmatrix}. \tag{E.6}$$

Another important effect of the surface is the breaking of the rotational invariance. In solution, only the difference $|\varphi_1 - \varphi_2|$ is of importance to characterize the structure. The presence of the surface introduces an anisotropy such that the individual values φ_1, φ_2 become important. The physical interpretation of this behavior is as follows: the fibrinogen dimer can approach the surface in different orientations, some of which might allow more conformations compatible with eq. (E.2). This was already discussed for the orientational screening in section 4.1. To characterize different orientations, the angle Φ_0 is introduced. Since experimentally the orientation of the adsorbed molecule can not be observed, it is prudent to average over Φ_0.

To satisfy the conditions in eq. (E.2) the allowed intervals for h_0, ϑ, γ_i, φ_i need to be restricted. This restricts the possible conformations that the fibrinogen model can take at the surface due to sterical hindrance. The allowed values for ϑ vary as a function of h_0. This is because too low or too high values of ϑ will lead to conformations in which the hinges are intersecting

the surface. The inequalities in eq. (E.2c) can be solved to give an appropriate interval for ϑ:

$$\sin\vartheta \in \left[-\frac{h_0 - R_C}{\xi}, \frac{h_0 - R_C}{\xi}\right] \tag{E.7}$$

This can not be reduced to $[0, \vartheta_{max}]$ exploiting the rotational symmetry of the model. The reason is that the hinges are not equivalent but have a dihedral shift of π and in general

$$P(\gamma, \varphi) \neq P(\gamma, \varphi + \pi)\,. \tag{E.8}$$

The positioning of the D-regions does not affect ϑ, because even at the boundaries of the interval in eq. (E.7) values for (γ_i, φ_i) can be found that are compatible with eq. (E.2), i.e. restrictions imposed by the D-region only become important when the hinges already collide with the surface. Depending on the exact distribution of (γ, φ) it might be that these angles are not allowed.

E.2 CALCULATING THE D-E-D ANGLE DISTRIBUTION

E.2.1 *Analytical Attempt*

After these consideration, one can try to reproduce the experimental distribution for the angle α as defined in Fig. 93. To calculate α using simple trigonometry, only the x- and y- coordinates of the two D-regions are needed:

$$\alpha = \cos^{-1}\left(\frac{r_{1,x}\, r_{2,x} + r_{1,y}\, r_{2,y}}{\sqrt{(r_{1,x}^2 + r_{1,y}^2)(r_{2,x}^2 + r_{2,y}^2)}}\right) \tag{E.9}$$

The usage of $\cos^{-1}(\ldots)$ implements the experimental bias that, because the connector regions are not visible, no values $\alpha > \pi$ can be measured.

The distribution $P(\alpha)$ could be constructed from the knowledge of its moments

$$\begin{aligned}\langle\alpha^n\rangle \;=\; & \int_0^H \frac{dh_0}{H}\int_0^{2\pi}\frac{d\Phi_0}{2\pi}\int_{-\vartheta_{max}(h_0)}^{\vartheta_{max}(h_0)}\frac{\xi d\vartheta}{2(h_0 - R_C)} \\ & \int_{\mathcal{C}_1} d\gamma_1\, d\varphi_1\, P_S(\gamma_1, \varphi_1 + \Phi_0) \\ & \int_{\mathcal{C}_2} d\gamma_2\, d\varphi_2\, P_S(\gamma_2, \varphi_2 + \pi + \Phi_0)\,\alpha^n\end{aligned} \tag{E.10}$$

via the moment generating function. To minimize any bias, uniform distributions h_0, ϑ, Φ_0 have been assumed. The areas $\mathcal{C}_{1/2}$ represent the regions in (γ, φ) space, that are compatible with eq. (E.2) using equations (E.5a) and (E.5b) respectively. It has also been included that the hinges are independent but rotated by π around the central region (see 2.1). The function $P_S(\gamma, \varphi)$ is the

distribution for the hinge angles that has been extracted from the simulation. That means all parts of the integral in eq. (E.10) are known and it can in principle be evaluated. Due to the complicated form of α, the areas $C_{1/2}$, and the only numerical knowledge of $P_S(\gamma, \varphi)$ it is not possible to perform this calculation analytically.

E.2.2 *The limit of a flat adsorbate*

Before the numerical solution is discussed, a few insights can be gained from further analytical considerations. A case which is of particular interest is the case of a flat adsorbate, i.e.

$$h_{max,D} \to 0,\ h_{max,E} \to 0\,. \tag{E.11}$$

This case allows one to asses to which part of the α-distribution the cis- and trans conformations on the surface contribute. For a comparison with experiments only quantities that can be measured on the projection of the fibrinogen structure onto the surface (see Fig. 155) are of interest.

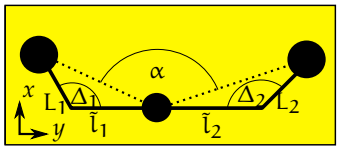

Figure 155: Top view of an adsorbed phenomenological model. All quantities are calculated from the projection of the molecule on the xy-plane.

In the flat adsorbate case, φ_i is determined by the solutions of

$$\sin(\varphi_i) = \frac{R_D - R_E \pm \cos(\gamma_i)\sin(\vartheta)}{\cos(\vartheta)\sin(\gamma_i)}\,. \tag{E.12}$$

It is clear that eq. (E.12) has no solutions for $\gamma \to 0$ because in this case the angle φ_i is ill defined. The projected bending angles at the hinges Δ_i can be determined as a function of γ_i, ϑ and which solution of eq. (E.12) is chosen. An explicit formula for Δ_i can be derived by using eq. (E.4) and eq. (E.6) but is not very illuminating. One can however determine that there has to be a minimum angle Δ_{min} because the D- and E-regions can not penetrate each other. To keep calculations simple, hard sphere interactions between the domains are assumed.

Let $\mathbf{a_i}$ be the projection of $\mathbf{r}_i$ on the xy-plane, then

$$\|\mathbf{a}_i\| \geqslant 2\sqrt{R_D R_E} \quad \text{(E.13a)} \qquad \|\mathbf{a}_1 - \mathbf{a}_2\| \geqslant 2R_D \quad \text{(E.13b)}$$

$$\mathbf{a}_i = \begin{pmatrix} L_i \sin(\Delta_i) \\ \pm(\tilde{l}_i - L_i \cos(\Delta_i)) \end{pmatrix} \tag{E.13c}$$

must hold. Here $\tilde{l}_i$ and L_i are defined as the projected connector length between the center of the E-region and hinge i and the connector between the center of D-region i and hinge i (see Fig. 155). The equality in eq. (E.13a) determines Δ_{min} as

$$\Delta_{min} = \cos^{-1}\left(\frac{L_i^2 + \tilde{l}_i^2 - 4R_D R_E}{2\tilde{l}_i L_i}\right) \tag{E.14}$$

and therefore also $\Delta_{max} = 2\pi - \Delta_{min}$. Equation (E.13b) on the other hand excludes regions where Δ_1, Δ_2 are both close to integer multiples of 2π. A comparison of equations (E.3) and (E.13c) shows that a similar minimum angle γ_{min} must exists for free fibrinogen .

The projected connector lengths L_i, $\tilde{l}_i$ are determined by ϑ:

$$\tilde{l}_i = \xi \cos(\vartheta) \tag{E.15a}$$

$$L_i = \sqrt{1 - (R_E - R_D \mp \sin(\vartheta))^2} \tag{E.15b}$$

Considering that $R_D \approx R_E$ and that, for the flat adsorbates discussed here, ϑ is limited to $[0, \frac{R_E}{\xi}]$, which corresponds to maximum angles of about 16°, eq. (E.15) can be approximated as:

$$\tilde{l}_i \approx \xi \tag{E.16a}$$

$$L_i \approx \sqrt{1 - (xR_E)^2} \approx 1 \quad x \in [0, 1] \tag{E.16b}$$

Because $R_E^2 \ll 1$, the tilting angle ϑ is unimportant for the calculation of

$$\alpha(\Delta_1, \Delta_2) = \cos^{-1}\left(\frac{\mathbf{a}_1 \cdot \mathbf{a}_2}{\|\mathbf{a}_1\| \|\mathbf{a}_2\|}\right) \tag{E.17}$$

in this limit.

The plots in Fig. 156 make it clear that the cis- as well as the trans-conformations can contribute to seemingly extended conformations with $\alpha \approx \pi$ but the trans-conformations do so over a much wider range of (Δ_1, Δ_2). It is also clear that there exists a minimum angle for both the cis- and the trans-configurations. These angles are

$$\alpha_{min}^{(trans)} = 77.5° \tag{E.18a}$$

$$\alpha_{min}^{(cis)} = 8.8° \tag{E.18b}$$

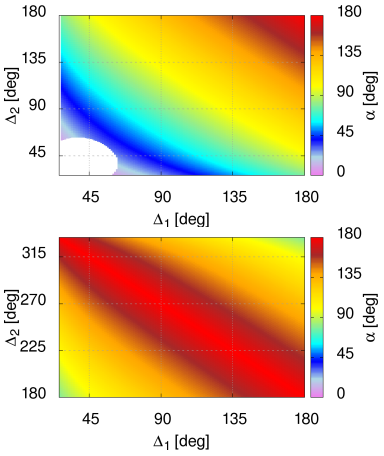

Figure 156: Values of the bending angle in the cis-conformation (top) and the trans-conformation (bottom). The white region in the lower left corner of the cis-region represents the region that is not allowed due to D-D collisions.

The minimum angles are determined by the conditions given in the equations (E.13a) and (E.13b), so different interactions between the domains could change them.

One would expect that most conformations have $90° < \Delta_i < 270°$ which would mean that the trans conformation only contributes to $\alpha > 120°$ and the cis conformation contributes to $\alpha > 80°$. Experiments do of course not fit this description neatly and even angles below 60° are measured. This means that in actual adsorbed fibrinogen both cis- and trans-conformations are likely present.

It is also interesting to note that in the case of a flat adsorbate with $R_D \neq R_E$ the hinge can not be unbent. An unbend hinge would require $\vartheta = 0$ as well as $\gamma_i = 0$, but in this case $h_{D,i} = R_E - R_D \neq 0$. That means if $\frac{R_D}{R_E}$ is such that the minimum bending angle does not correspond to a minimal energetic

state this could drive mechanisms not included in the model, like denaturation of the globular regions or a breakdown of the stiffness assumption of the connectors.

To simplify things further, one can now additionally look at the limits $R_D \to 0$, $R_E \to 0$. In this case, fibrinogen conformations correspond to 5 non-interacting points and $\vartheta = 0$ exactly. According to eq. (E.12), φ_i can in this case be either 0 or π. Due to the π shift between the dihedral angles in the protomers and the fact that $\gamma_i < \pi$, choosing $\varphi_1 \neq \varphi_2$ corresponds to the cis conformation while $\varphi_1 = \varphi_2$ corresponds to the trans conformation. The projection of eq. (E.4) into the xy-plane then becomes

$$\begin{pmatrix} r_{i,x} \\ r_{i,y} \end{pmatrix} = \begin{pmatrix} \sin(\gamma)\cos(\varphi_i) \\ \pm(\xi - \cos(\gamma_i)) \end{pmatrix} \tag{E.19}$$

From which it is clear that in this case $\Delta_i = \gamma_i$ and no Δ_{min} exists. The additionally available regions in the γ_1, γ_2 plane contribute mostly to conformations with small α. This is not surprising considering that the two D-regions end up in the same half plane $y > 0$ or $y < 0$.

E.2.3 *First Numerical Attempt*

To get less aproximate results, a numerical calculation of $P(\alpha)$ has to be performed. An attempt can be made based on the moment generating functions presented above. One can rewrite the integration over $\mathcal{C}_{1/2}$ in eq. (E.10) to an integral over the whole (γ, φ) plane if one introduces an appropriate function which is 1 on $\mathcal{C}_{1/2}$ and 0 otherwise. A simple way to produce such a function would be by multiplying two step functions:

$$S_{\mathcal{C}}(\gamma, \varphi) = \Theta(\varphi - f_l(\gamma))\Theta(f_u(\gamma) - \varphi)\,, \tag{E.20}$$

where the functions describing the upper and lower bound of C are determined by solving:

$$h_{D,1/2}(h_0, \vartheta, \gamma, f_l(\gamma)) = 0 \tag{E.21a}$$
$$h_{D,1/2}(h_0, \vartheta, \gamma, f_u(\gamma)) = h_{max,D} \tag{E.21b}$$

Because the functions $h_{D,i}$ depend on ϑ and h_0, $S_{\mathcal{C}} = S_{\mathcal{C},\vartheta,h_0}$ will inherit a parametric dependence on these parameters. The same applies if the roles of γ and φ are exchanged. In general, the functions f_u and f_l can only be determined numerically. Even this, is not without its problems, because the equations are nonlinear and typically more than one solution exists. Nevertheless, the integral in eq. (E.10) can be written as

$$\langle \alpha^n \rangle = \int d^7\underline{F}\,\widetilde{P}(\underline{F})\,\alpha^n(\underline{F}),\ \underline{F} = (h_0, \Phi_0, \vartheta, \gamma_1, \gamma_2, \varphi_1, \varphi_2)\,, \tag{E.22}$$

where $\widetilde{P}(\underline{F})$ represents the probability density of finding a specific fibrinogen conformation $\underline{F}$. Knowing $\widetilde{P}(\underline{F})$ provides a great deal of information. Indeed, it provides all the information necessary to characterize the adsorbed states of the phenomenological model in the infinite dilution limit. For example, the distribution $P(\varphi_1, \varphi_2)$ could be calculated by integrating out the other degrees of freedom. This would be an alternate way to calculate the frequency with wich cis- and trans- conformations occur.

A more successful numerical approach is the Monte Carlo procedure discussed in Chapter 8.

E.3 RESULTS OF THE FITTING

In the following the results for a_{ij} of the least square fitting are presented for the mica and graphite surface. The numbering of the regions is shown in Fig. 158. The bending angle distributions resulting from restricting the hinges to specific regions are shown in Fig. 157. From these plots it is clear that distributions $P_{ii}(\alpha)$ contribute siginificantly only to peaks at $\gamma = 180°$.

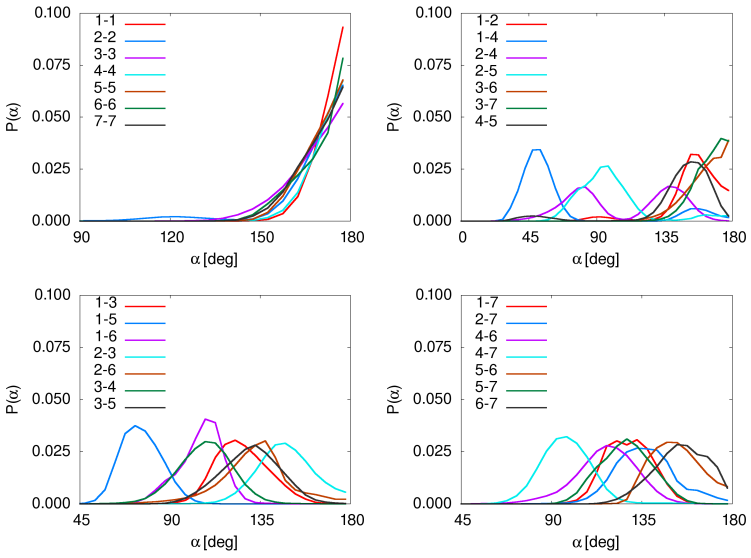

Figure 157: Bending angle distribution $P_{ij}(\alpha)$ (see Fig. 93) if the hinges are restricted to different regions in the $P(\gamma, \varphi)$ distribution (see Fig. 158).

The regions 2,6, and 7 in Fig. 158 represent transition regions and are not expected to give major contributions on experimental time scales. Therefore they are neglected. In the independent case only the four parameters a_{ii} are necessary to fit the distribution. The results for these parameters are given in Table 30. In the correlated case 10 independent values have to be fitted to specify the symmetric matrix a_{ij}. These values are given in Table 31 and Tables 32. The resulting distributions are shown in Fig. 95 in Chapter 8.

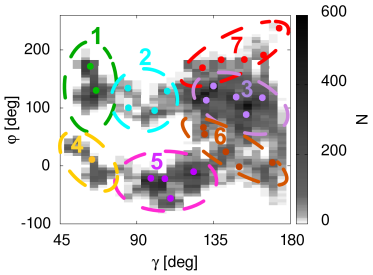

Figure 158: Numbering of the division of $P(\gamma, \phi)$ in Fig. 94

i	a_i	
	Graphite	*Mica*
1	*0.000*	*0.121*
3	*0.225*	*0.327*
4	*0.164*	*0.244*
5	*0.611*	*0.308*

Table 30: Results for the fitted parameters of the phenomenological model with uncorrelated hinges. The model was fitted to experimental results on mica [17] and hydrophilic graphite [143].

	1	3	4	5
1	0.018			
3	0.000	0.070		
4	0.000	0.000	0.016	
5	0.000	0.214	0.176	0.506

Table 31: Results for the fitted parameters of the phenomenological model with correlated hinges. The model was fitted to experimental results on hydrophilic graphite [143]

	1	3	4	5
1	0.001			
3	0.001	0.161		
4	0.000	0.119	0.000	
5	0.048	0.131	0.539	0.001

Table 32: Results for the fitted parameters of the phenomenological model with correlated hinges. The model was fitted to experimental results on mica [17]

LIST OF COMMON SYMBOLS AND ABBREVIATIONS

The following is a reference list of symbols and abbreviations that occur frequently throughout the thesis. Symbols that appear only in single calculation are not listed.

AFM	Atomic Force Microscope
CG	Coarse Graining
DRMS	Distance Root Mean Square Deviation
ED	Essential Dynamics
EDCG	Essential Dynamics Coarse graining
ENM	Elastic Network model
Fg	Fibrinogen
Fp	Fibrinopeptide
hENM	heterogeneous Elastic Network Model
LMI	Linear Mutual Information
PCA	Principle Component Analysis
PDB	Protein Data Bank
PDBID	Protein Data Bank Identification Code
P1,P2	Integrin binding sites in the fibrinogen γC domain, see Table 2
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
RSA	Random Sequential Adsorption

$\Phi(\mathbf{r})$ Electrostatic Potential

(γ, φ) hinge bending and dihedral angle

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

[1] Weber, G. In *Advances in Protein Chemistry*, C.B. Anfinsen, J. T. E. and Richards, F. M., editors, volume 29 of *Advances in Protein Chemistry*, 1–83. Academic Press (1975).

[2] Weisel, J. W. In *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*, David A. D. Parry and Squire, J. M., editors, volume 70 of *Advances in Protein Chemistry*, 247–299. Academic Press (2005).

[3] Voet, D. and Voet, J. G. *Biochemistry*. Wiley, 3rd edition, (2004).

[4] www.pdb.org.

[5] Wormald, M. R. and Dwek, R. A. *Structure* **7**(7), R155 – R160 (1999).

[6] Ellis, C. R., Maiti, B., and Noid, W. G. *Journal of the American Chemical Society* **134**(19), 8184–8193 (2012).

[7] Jo, S., Lee, H. S., Skolnick, J., and Im, W. *PLoS Computational Biology* **9**(3), e1002946 (2013).

[8] Levitt, M. and Warshel, A. *Nature* **253**(5494), 694–698 (1975).

[9] Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. *Science* **334**(6055), 517–520 (2011).

[10] Dill, K. A. and MacCallum, J. L. *Science* **338**(6110), 1042–1046 (2012).

[11] Kiss, G., Celebi-Ölcüm, N., Moretti, R., Baker, D., and Houk, K. N. *Angewandte Chemie International Edition* **52**(22), 5700–5725 (2013).

[12] Iben, I. E. T., Braunstein, D., Doster, W., Frauenfelder, H., Hong, M. K., Johnson, J. B., Luck, S., Ormos, P., Schulte, A., Steinbach, P. J., Xie, A. H., and Young, R. D. *Physical Review Letters* **62**, 1916–1919 (1989).

[13] Frauenfelder, H., Sligar, S., and Wolynes, P. *Science* **254**(5038), 1598–1603 (1991).

[14] Doolittle, R. F. *Annual Review Biochemistry* **64**, 287–314 (1995).

[15] Lupas, A. N. and Gruber, M. *Advances in Protein Chemistry* **70**, 37–78 (2005).

[16] Parry, D. A. D. *Advances in Protein Chemistry* **70**, 11–35 (2005).

[17] Yermolenko, I. S., Lishko, V. K., Ugarova, T. P., and Magonov, S. N. *Biomacromolecules* **12**(2), 370–379 (2011).

[18] Hall, K., C. E. and Slayter, H. S. *The Journal of Biophysical and Biochemical Cytology* **5**(1), 11–16 (1959).

[19] Kollman, J., Pandi, L., Sawaya, M., Riley, M., and Doolittle, R. *Biochemistry* **48**(18), 3877–3886 (2009).

[20] Doolittle, R. F. and Kollman, J. M. *Proteins* **63**(2), 391–397 (2006).

[21] Burton, R. A., Tsurupa, G., Medved, L., and Tjandra, N. *Biochemistry* **45**(7), 2257–2266 (2006).

[22] Burton, R. A., Tsurupa, G., Hantgan, R. R., Tjandra, N., and Medved, L. *Biochemistry* **46**(29), 8550–8560 (2007).

[23] Medved, L., Litvinovich, S., Ugarova, T., Matsuka, Y., and Ingham, K. *Biochemistry* **36**(15), 4685–4693 (1997).

[24] Everse, S. J., Spraggon, G., Veerapandian, L., Riley, M., and Doolittle, R. F. *Biochemistry* **37**(24), 8637–8642 (1998).

[25] Russell F. Doolittle, Danielle M. Goldbaum, L. R. D. *Journal of Molecular Biology* **120**(2), 311 (1978).

[26] Veklich, Y. I., Gorkun, O. V., Medved, L. V., Nieuwenhuizen, W., and Weisel, J. W. *Journal of Biological Chemistry* **268**(18), 13577–13585 (1993).

[27] Adamczyk, Z., Cichocki, B., Ekiel-Jezewska, M. L., Slowicka, A., Wajnryb, E., and Wasilewska, M. *Journal of Colloid and Interface Science* **385**(1), 244 – 257 (2012).

[28] Madrazo, J., Brown, J. H., Litvinovich, S., Dominguez, R., Yakovlev, S., Medved, L., and Cohen, C. *Proceedings of the National Academy of Sciences of the United States of America* **98**(21), 11967–11972 (2001).

[29] Spraggon, G., Everse, S. J., and Doolittle, R. F. *Nature* **389**, 455 (1997).

[30] Al-Mondhiry, H. and Ehmann, W. C. *American Journal of Hematology* **46**, 343 (1994).

[31] Doolittle, R. F. *The evolution of vertebrate blood clotting*. University Science Books, (2012).

[32] Takagi, T. and Doolittle, R. *Thrombosis Research* **7**(5), 813 – 818 (1975).

[33] Mihalyi, E., Weinberg, R. M., Towne, D. W., and Friedman, M. E. *Biochemistry* **15**(24), 5372–5381 (1976).

[34] Shrager, R. I., Mihalyi, E., and Towne, D. W. *Biochemistry* **15**(24), 5382–5386 (1976).

[35] Mihalyi, E. *Annals of the New York Academy of Sciences* **408**(1), 60–70 (1983).

[36] Takagi, T. and Doolittle, R. F. *Biochemistry* **14**(5), 940–946 (1975).

[37] Marsh, J. J., Guan, H. S., Li, S., Chiles, P. G., Tran, D., and Morris, T. A. *Biochemistry* **52**(32), 5491–5502 (2013).

[38] Walker, J. B. and Nesheim, M. E. *The Journal of Biological Chemistry* **274**(8), 5201–5212 (1999).

[39] Reinhart, W. H. *Vascular Medicine* **8**(3), 211–216 (2003).

[40] Yakubenko, V. P., Solovjov, D. A., Zhang, L., Yee, V. C., Plow, E. F., and Ugarova, T. P. *Journal of Biological Chemistry* **276**(17), 13995–14003 (2001).

[41] Lishko, V. K., Podolnikova, N. P., Yakubenko, V. P., Yakovlev, S., Medved, L., Yadav, S. P., and Ugarova, T. P. *Journal of Biological Chemistry* **279**(43), 44897–44906 (2004).

[42] Podolnikova, N. P., Yakubenko, V. P., Volkov, G. L., Plow, E. F., and Ugarova, T. P. *Journal of Biological Chemistry* **278**(34), 32251–32258 (2003).

[43] Podolnikova, N. P., Gorkun, O. V., Loreth, R. M., Yee, V. C., Lord, S. T., and Ugarova, T. P. *Biochemistry* **44**(51), 16920–16930 (2005).

[44] Mosesson, M. W. *Journal of Thrombosis and Haemostasis* **3**(8), 1894–1904 (2005).

[45] Vogel, V. and Sheetz, M. *Nature Reviews Molecular Cell Biology* **7**(4), 265–275 (2006).

[46] Podolnikova, N. P., Yermolenko, I. S., Fuhrmann, A., Lishko, V. K., Magonov, S., Bowen, B., Enderlein, J., Podolnikov, A. V., Ros, R., and Ugarova, T. P. *Biochemistry* **49**(1), 68–77 (2010).

[47] Kotlín, R., Reicheltová, Z., Malý, M., Suttnar, J., Sobotková, A., Salaj, P., Hirmerová, J., Riedel, T., and Dyr, J. E. *Thrombosis and Haemostasis* **102**(3), 479–486 (2009).

[48] Weisel, J. W. *Biophysical Chemistry* **112**(2–3), 267 (2004).

[49] Brennan, S. O., Davis, R. L., Lowen, R., and Ruskova, A. *Haematologica* **94**(4), 585–588 (2009).

[50] da Costa, G., Gomes, R. A., Guerreiro, A., Mateus, Ã., Monteiro, E., Barroso, E., Coelho, A. V., Freire, A. P., and Cordeiro, C. *PLoS ONE* **6**(10), e24850 (2011).

[51] Uemichi, T., Liepnieks, J., Gertz, M., and Benson, M. *Amyloid* **5**(3), 188–192 (1998).

[52] Yakovlev, S., Gorlatov, S., Ingham, K., and Medved, L. *Biochemistry* **42**(25), 7709–7716 (2003).

[53] Makogonenko, E., Tsurupa, G., Ingham, K., and Medved, L. *Biochemistry* **41**(25), 7907–7913 (2002).

[54] Macheboeuf, P., Buffalo, C., yu Fu, C., Zinkernagel, A. S., Cole, J. N., Johnson, J. E., Nizet, V., and Ghosh, P. *Nature* **472**, 64 (2011).

[55] Laudano, A. P. and Doolittle, R. F. *Proceedings of the National Academy of Sciences of the United States of America* **75**(7), 3085–3089 (1978).

[56] Doolittle, R. F., Chen, A., and Pandi, L. *Biochemistry* **45**(47), 13962–13969 (2006).

[57] Geer, C. B., Tripathy, A., Schoenfisch, M. H., Lord, S. T., and Gorkun, O. V. *Journal of Thrombosis and Haemostasis* **5**(12), 2344–2351 (2007).

[58] Weisel, J. W., Nagaswami, C., and Makowski, L. *Proceedings of the National Academy of Sciences of the United States of America* **84**(24), 8991–8995 (1987).

[59] Brzoska, T., Suzuki, Y., Mogami, H., Sano, H., and Urano, T. *PLoS ONE* **8**(2), e55466 (2013).

[60] Weisel, J. W. *Biophysical Journal* **50**(6), 1079–1093 (1986).

[61] Pechik, I., Yakovlev, S., Mosesson, M. W., Gilliland, G. L., and Medved, L. *Biochemistry* **45**(11), 3588–3597 (2006).

[62] Yang, Z., Mochalkin, I., and Doolittle, R. F. *Proceedings of the National Academy of Sciences of the United States of America* **97**(26), 14156–14161 (2000).

[63] Fowler, W. E., Hantgan, R. R., Hermans, J., and Erickson, H. P. *Proceedings of the National Academy of Sciences of the United States of America* **78**(8), 4872–4876 (1981).

[64] Protopopova, A. D., Barinov, N. A., Zavyalova, E. G., Kopylov, A. M., Sergienko, V. I., and Klinov, D. V. *Journal of Thrombosis and Haemostasis* **13**(4), 570–579 (2015).

[65] Tsurupa, G., Pechik, I., Litvinov, R. I., Hantgan, R. R., Tjandra, N., Weisel, J. W., and Medved, L. *Biochemistry* **51**(12), 2526–2538 (2012).

[66] Weisel, J. W. *Science* **320**(5875), 456–457 (2008).

[67] Blombäck, B. and Bark, N. *Biophysical Chemistry* **112**(2–3), 147 – 151 (2004).

[68] Helms, C. C., Ariëns, R. A. S., Uitte de Willige, S., Standeven, K. F., and Guthold, M. *Biophysical Journal* **102**(1), 168–175 (2012).

[69] Doolittle, R. F., Jiang, Y., and Nand, J. *Journal of Molecular Evolution* **66**(2), 185–196 (2008).

[70] Doolittle, R. F. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 35–40 (2009).

[71] Ponczek, M. B., Gailani, D., and Doolittle, R. F. *Journal of Thrombosis and Haemostasis* **6**(11), 1876–1883 (2008).

[72] Doolittle, R. F., McNamara, K., and Lin, K. *Protein Science* **21**(12), 1808–1823 (2012).

[73] Kairies, N., Beisel, H. G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S., Bode, W., Huber, R., and Kawabata, S. *Proceedings of the National Academy of Sciences of the United States of America* **98**(24), 13519–13524 (2001).

[74] Doolittle, R. F. *Annals of the New York Academy of Sciences* **408**, 13–27 (1983).

[75] Nakanishi, K., Sakiyama, T., and Imamura, K. *Journal of Bioscience and Bioengineering* **91**(3), 233–244 (2001).

[76] Kasemo, B. *Surface Science* **500**(1–3), 656–677 (2002).

[77] Engelhardt, K., Rumpel, A., Walter, J., Dombrowski, J., Kulozik, U., Braunschweig, B., and Peukert, W. *Langmuir* **28**(20), 7780–7787 (2012).

[78] Alharthi, S. A., Benavidez, T. E., and Garcia, C. D. *Langmuir* **29**(10), 3320–3327 (2013).

[79] Soderquist, M. and Walton, A. *Journal of Colloid and Interface Science* **75**(2), 386 – 397 (1980).

[80] Devineau, S., Zanotti, J.-M., Loupiac, C., Zargarian, L., Neiers, F., Pin, S., and Renault, J. P. *Langmuir* **29**(44), 13465–13472 (2013).

[81] Calcaterra, J., Van Cott, K. E., Butler, S. P., Gil, G. C., Germano, M., van Veen, H. A., Nelson, K., Forsberg, E. J., Carlson, M. A., and Velander, W. H. *Biomacromolecules* **14**(1), 169–178 (2013).

[82] Wnek, G. E., Carr, M. E., Simpson, D. G., and Bowlin, G. L. *Nano Letters* **3**(2), 213–216 (2003).

[83] Martino, M. M., Briquez, P. S., Ranga, A., Lutolf, M. P., and Hubbell, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **110**(12), 4563–4568 (2013).

[84] Kim, B.-S., Kim, J. S., and Lee, J. *Journal of Biomedical Materials Research Part A* **101A**(9), 2661–2666 (2013).

[85] Salem, S. A., Hwei, N. M., Saim, A. B., Ho, C. C., Sagap, I., Singh, R., Yusof, M. R., Md Zainuddin, Z., and Bt. Hj Idrus, R. *Journal of Biomedical Materials Research Part A* **101A**(8), 2237–2247 (2013).

[86] Brown, E. E., Hu, D., Abu Lail, N., and Zhang, X. *Biomacromolecules* **14**(4), 1063–1071 (2013).

[87] Lalani, R. and Liu, L. *Biomacromolecules* **13**(6), 1853–1863 (2012).

[88] Li, J., Zhang, Y., Yang, J., Tan, H., Li, J., and Fu, Q. *Journal of Biomedical Materials Research Part A* **101A**(5), 1362–1372 (2012).

[89] Nune, C., Xu, W., and Misra, R. D. K. *Journal of Biomedical Materials Research Part A* **100A**(12), 3197–3204 (2012).

[90] Donahoe, C. D., Cohen, T. L., Li, W., Nguyen, P. K., Fortner, J. D., Mitra, R. D., and Elbert, D. L. *Langmuir* **29**(12), 4128–4139 (2013).

[91] Baier, R. E. and Dutton, R. C. *Journal of Biomedical Materials Research* **3**(1), 191–206 (1969).

[92] Grunkemeier, J., Tsai, W., McFarland, C., and Horbett, T. *Biomaterials* **21**(22), 2243 – 2252 (2000).

[93] Lishko, V. K., Burke, T., and Ugarova, T. *Blood* **109**(4), 1541–1549 (2007).

[94] Tang, L. and Eaton, J. W. *The Journal of Experimental Medicine* **178**(6), 2147–2156 (1993).

[95] Gupta, V., Aravamuthan, B. R., Baskerville, S., Smith, S. K., Gupta, V., Lauer, M. A., and Fischell, T. A. *Journal of Invasive Cardiology* **16**(6), 304–310 (2004).

[96] Tang, L., Ugarova, T. P., Plow, E. F., and Eaton, J. W. *The Journal of Clinical Investigation* **97**(5), 1329–1334 (1996).

[97] Kuo, W.-H., Wang, M.-J., Chien, H.-W., Wei, T.-C., Lee, C., and Tsai, W.-B. *Biomacromolecules* **12**(12), 4348–4356 (2011).

[98] Corum, L. E., Eichinger, C. D., Hsiao, T. W., and Hlady, V. *Langmuir* **27**(13), 8316–8322 (2011).

[99] Vroman, L. *Nature* **196**, 476–477 Nov (1962).

[100] Vroman, L., Adams, A., Fischer, G., and Munoz, P. *Blood* **55**(1), 156–159 (1980).

[101] Lesniak, A., Fenaroli, F., Monopoli, M. P., Aberg, C., Dawson, K. A., and Salvati, A. *ACS Nano* **6**(7), 5845–5857 (2012).

[102] Jedlovszky-Hajdu, A., Bombelli, F. B., Monopoli, M. P., Tombacz, E., and Dawson, K. A. *Langmuir* **28**(42), 14983–14991 (2012).

[103] de Mel, A., Chaloupka, K., Malam, Y., Darbyshire, A., Cousins, B., and Seifalian, A. M. *Journal of Biomedical Materials Research Part A* **100A**, 2348–2357 (2012).

[104] Deng, Z. J., Liang, M., Toth, I., Monteiro, M. J., and Minchin, R. F. *ACS Nano* **6**(10), 8962–8969 (2012).

[105] Hantgan, R. R. *Biochemistry* **21**(8), 1821–1829 (1982).

[106] Acuna, A., Gonzalez-Rodriguez, J., Lillo, M., and Naqvi, K. *Biophysical Chemistry* **26**(1), 55 – 61 (1987).

[107] Adamczyk, Z., Nattich, M., and Wasilewska, M. *Adsorption* **16**, 259–269 (2010).

[108] Yermolenko, I. S., Fuhrmann, A., Magonov, S. N., Lishko, V. K., Oshkadyerov, S. P., Ros, R., and Ugarova, T. P. *Langmuir* **26**(22), 17269–17277 (2010).

[109] Lishko, V. K., Yermolenko, I. S., and Ugarova, T. P. *Journal of Thrombosis and Haemostasis* **8**(4), 799–807 (2010).

[110] Shen, L., Adachi, T., Vanden Bout, D., and Zhu, X.-Y. *Journal of the American Chemical Society* **134**(34), 14172–14178 (2012).

[111] Stavis, C., Clare, T. L., Butler, J. E., Radadia, A. D., Carr, R., Zeng, H., King, W. P., Carlisle, J. A., Aksimentiev, A., Bashir, R., and Hamers, R. J. *Proceedings of the National Academy of Sciences of the United States of America* **108**(3), 983–988 (2011).

[112] Clarke, M. L., Wang, J., and Chen, Z. *The Journal of Physical Chemistry B* **109**(46), 22027–22035 (2005).

[113] Wasilewska, M. and Adamczyk, Z. *Langmuir* **27**(2), 686–696 (2011).

[114] Adamczyk, Z., Bratek-Skicki, A., Dabrowska, P., and Nattich-Rak, M. *Langmuir* **28**(1), 474–485 (2012).

[115] Ortega-Vinuesa, J., Tengvall, P., and Lundström, I. *Thin Solid Films* **324**(1–2), 257 – 273 (1998).

[116] Tsapikouni, T. S. and Missirlis, Y. F. *Colloids and Surfaces B: Biointerfaces* **57**(1), 89 – 96 (2007).

[117] Pandey, L. M., Pattanayek, S. K., and Delabouglise, D. *The Journal of Physical Chemistry C* **117**(12), 6151–6160 (2013).

[118] Bratek-Skicki, A., Zeliszewska, P., Adamczyk, Z., and Ciesla, M. *Langmuir* **29**(11), 3700–3710 (2013).

[119] Yang, Z.-L., Zhou, S., Lu, L., Wang, X., Wang, J., and Huang, N. *Journal of Biomedical Materials Research Part A* **100A**, 3124–3133 (2012).

[120] Malmsten, Muller, and Lassen. *Journal of Colloid and Interface Science* **193**(1), 88–95 (1997).

[121] Lee, W.-H., Zavgorodniy, A. V., Loo, C.-Y., and Rohanizadeh, R. *Journal of Biomedical Materials Research Part A* **100A**(6), 1539–1549 (2012).

[122] Dolatshahi-Pirouz, A., Foss, M., and Besenbacher, F. *The Journal of Physical Chemistry C* **115**(28), 13617–13623 (2011).

[123] Sommerfeld, J., Richter, J., Niepelt, R., Kosan, S., Keller, T., Jandt, K., and Ronning, C. *Biointerphases* **7**, 1–7 (2012).

[124] Van De Keere, I., Willaert, R., Hubin, A., and Vereecken, J. *Langmuir* **24**(5), 1844–1852 (2008).

[125] Hemmerlé, J., Altmann, S. M., Maaloum, M., Hörber, J. K. H., Heinrich, L., Voegel, J.-C., and Schaaf, P. *Proceedings of the National Academy of Sciences of the United States of America* **96**(12), 6705–6710 (1999).

[126] Sit, P. S. and Marchant, R. E. *Thrombosis and Haemostasis* **82**, 1005 (1999).

[127] Agnihotri, A. and Siedlecki, C. A. *Langmuir* **20**(20), 8846–8852 (2004).

[128] Fernandez-Montes Moraleda, B., Roman, J. S., and Rodriguez-Lorenzo, L. M. *Journal of Biomedical Materials Research Part A* **101A**(8), 2332–2339 (2013).

[129] Tunc, S., Maitz, M. F., Steiner, G., Vazquez, L., Pham, M. T., and Salzer, R. *Colloids and Surfaces B: Biointerfaces* **42**(3–4), 219 – 225 (2005).

[130] Soman, P., Rice, Z., and Siedlecki, C. A. *Langmuir* **24**(16), 8801–8806 (2008).

[131] Deng, Z. J., Liang, M., Monteiro, M., Toth, I., and Minchin, R. F. *Nature Nanotechnology* **6**(1), 39–44 (2011).

[132] Sit, P. and Marchant, R. E. *Surface Science* **491**(3), 421 – 432 (2001).

[133] Riedel, T., Suttnar, J., Brynda, E., Houska, M., Medved, L., and Dyr, J. E. *Blood* **117**(5), 1700–1706 (2011).

[134] Dyr, J., Tichy, I., Jirouskova, M., Tobiska, P., Slavik, R., Homola, J., Brynda, E., Houska, M., and Suttnar, J. *Sensors and Actuators B: Chemical* **51**(1–3), 268 – 272 (1998).

[135] Lewis, K. B. and Ratner, B. D. *Colloids and Surfaces B: Biointerfaces* **7**(5–6), 259 – 269 (1996).

[136] Wigren, R., Elwing, H., Erlandsson, R., Welin, S., and Lundström, I. *FEBS Letters* **280**(2), 225–228 (1991).

[137] Williams, R. C. *Journal of Molecular Biology* **150**(3), 399–408 (1981).

[138] Zavyalova, E. G., Protopopova, A. D., Kopylov, A. M., and Yaminsky, I. V. *Langmuir* **27**(8), 4922–4927 (2011).

[139] Cacciafesta, P., Humphris, A. D. L., Jandt, K. D., and Miles, M. J. *Langmuir* **16**(21), 8167–8175 (2000).

[140] Katan, A. J. and Dekker, C. *Cell* **147**, 979 (2011).

[141] Taatjes, D., Quinn, A., Jenny, R., P., H., Bovill, E., and McDonagh, J. *Cell Biology International* **21**(11), 715–726 (1997).

[142] Marchin, K. L. and Berrie, C. L. *Langmuir* **19**(23), 9883–9888 (2003).

[143] Beijbom, L., Larsson, U., Kaveus, U., and Hebert, H. *Journal of Ultrastructure and Molecular Structure Research* **98**(3), 312 – 319 (1988).

[144] Koo, J., Rafailovich, M. H., Medved, L., Tsurupa, G., Kudryk, B. J., Liu, Y., and Galanakis, D. K. *Journal of Thrombosis and Haemostasis* **8**(12), 2727–2735 (2010).

[145] Koo, J., Galanakis, D., Liu, Y., Ramek, A., Fields, A., Ba, X., Simon, M., and Rafailovich, M. H. *Biomacromolecules* **13**, 1259 (2012).

[146] Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S., and Xiao, X. *Quarterly Reviews of Biophysics* **44**(2), 191–228 (2011).

[147] Binder, K. and Heermann, D. W. *Monte Carlo Simulation in Statistical Physics*. Solid State Science. Springer, (2002).

[148] Frenkel, D. and Smit, B. *Understanding Molecular Simulations*, volume 1 of *Computational Science*. Academic Press, 2nd edition, (2002).

[149] Gronbech-Jensen, N., Hayre, N. R., and Farago, O. *Computer Physics Communications* **185**(2), 524–527 (2014).

[150] Gronbech-Jensen, N. and Farago, O. *The Journal of Chemical Physics* **141**(19), 194108 (2014).

[151] Iype, E., Hütter, M., Jansen, A. P. J., Nedea, S. V., and Rindt, C. C. M. *Journal of Computational Chemistry* **34**(13), 1143–1154 (2013).

[152] Ponder, J. W. and Case, D. A. In *Protein Simulations*, Daggett, V., editor, volume 66 of *Advances in Protein Chemistry*, 27–85. Academic Press (2003).

[153] Chang, L., Ishikawa, T., Kuwata, K., and Takada, S. *Journal of Computational Chemistry* **34**(14), 1251–1257 (2013).

[154] Ho, T. A. and Striolo, A. *The Journal of Chemical Physics* **138**(5), 054117 (2013).

[155] Tzanov, A. T., Cuendet, M. A., and Tuckerman, M. E. *The Journal of Physical Chemistry B* **118**(24), 6539–6552 (2014).

[156] Beauchamp, K. A., Lin, Y.-S., Das, R., and Pande, V. S. *Journal of Chemical Theory and Computation* **8**(4), 1409–1414 (2012).

[157] Kulik, H. J., Luehr, N., Ufimtsev, I. S., and Martinez, T. J. *The Journal of Physical Chemistry B* **116**(41), 12501–12509 (2012).

[158] Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and van Gunsteren, W. F. *European Biophysics Journal* **40**(7), 843–856 (2011).

[159] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *Journal of the American Chemical Society* **117**(19), 5179–5197 (1995).

[160] Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. *Journal of the American Chemical Society* **118**(45), 11225–11236 (1996).

[161] MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. *The Journal of Physical Chemistry B* **102**(18), 3586–3616 (1998).

[162] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. *The Journal of Chemical Physics* **79**(2), 926–935 (1983).

[163] Mackerell, A. D., Feig, M., and Brooks, C. L. *Journal of Computational Chemistry* **25**(11), 1400–1415 (2004).

[164] Piana, S., Lindorff-Larsen, K., and Shaw, D. E. *Biophysical Journal* **100**(9), L47–L49 (2011).

[165] Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and Mackerell, Jr, A. D. *Journal of Chemical Theory and Computation* **8**(9), 3257–3273 (2012).

[166] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. *Journal of Computational Chemistry* **31**(4), 671–690 (2010).

[167] Guvench, O., Mallajosyula, S. S., Raman, E. P., Hatcher, E., Vanommeslaeghe, K., Foster, T. J., Jamison, F. W., and MacKerell, A. D. *Journal of Chemical Theory and Computation* **7**(10), 3162–3180 (2011).

[168] Patel, D. S., Pendrill, R., Mallajosyula, S. S., Widmalm, G., and MacKerell, A. D. *The Journal of Physical Chemistry B* **118**(11), 2851–2871 (2014).

[169] Cruz-Chu, E. R., Aksimentiev, A., and Schulten, K. *The Journal of Physical Chemistry B* **110**(43), 21497–21508 (2006).

[170] Heinz, H., Koerner, H., Anderson, K. L., Vaia, R. A., and Farmer, B. L. *Chemistry of Materials* **17**(23), 5658–5669 (2005).

[171] Heinz, H., Vaia, R. A., Farmer, B. L., and Naik, R. R. *The Journal of Physical Chemistry C* **112**(44), 17281–17290 (2008).

[172] Herbers, C. R., Li, C., and van der Vegt, N. F. A. *Journal of Computational Chemistry* (2013).

[173] Latour, R. *Biointerphases* **3**, FC2–FC12 (2008).

[174] Bacteria and anatomy image adapted from public domain images from WikiMedia Commons:
http://commons.wikimedia.org/wiki/File:Laskowski_Anatomie_normale_05.jpg
http://commons.wikimedia.org/wiki/File:Average_prokaryote_cell-_en.svg.

[175] Warshel, A. and Levitt, M. *Journal of Molecular Biology* **103**(2), 227–249 (1976).

[176] Warshel, A. and Karplus, M. *Journal of the American Chemical Society* **96**(18), 5677–5689 (1974).

[177] Kong, Y. and Karplus, M. *Structure* **15**(5), 611–623 (2007).

[178] Balamurugan, D., Aquino, A. J. A., de Dios, F., Flores, L., Lischka, H., and Cheung, M. S. *The Journal of Physical Chemistry B* **117**(40), 12065–12075 (2013).

[179] The Royal Swedish Academy of Sciences. *Development of multiscale models for complex chemical systems*, (2013).

[180] Murtola, T., Bunker, A., Vattulainen, I., Deserno, M., and Karttunen, M. *Physical Chemistry Chemical Physics* **11**, 1869–1892 (2009).

[181] Praprotnik, M., Site, L. D., and Kremer, K. *Annual Review of Physical Chemistry* **59**(1), 545–571 (2008).

[182] Saunders, M. G. and Voth, G. A. *Annual Review of Biophysics* **42**(1), 73–93 (2013).

[183] Saunders, M. G. and Voth, G. A. *Current Opinion in Structural Biology* **22**(2), 144 – 150 (2012).

[184] Tozzini, V. *Current Opinion in Structural Biology* **15**(2), 144–150 (2005).

[185] Tozzini, V. *Quarterly Reviews of Biophysics* **43**, 333–371 (2010).

[186] Peter, C. and Kremer, K. *Faraday Discussions* **144**, 9–24 (2010).

[187] Benson, N. C. and Daggett, V. *The Journal of Physical Chemistry B* **116**(29), 8722–8731 (2012).

[188] Senn, H. M. and Thiel, W. *Angewandte Chemie International Edition* **48**(7), 1198–1229 (2009).

[189] Marrink, S. J., de Vries, A. H., and Mark, A. E. *The Journal of Physical Chemistry B* **108**(2), 750–760 (2004).

[190] Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. *Journal of Chemical Theory and Computation* **4**(5), 819–834 (2008).

[191] Bereau, T., Globisch, C., Deserno, M., and Peter, C. *Journal of Chemical Theory and Computation* **8**(10), 3750–3758 (2012).

[192] Gohlke, H. and Thorpe, M. *Biophysical Journal* **91**(6), 2115–2120 (2006).

[193] Zhang, Z., Lu, L., Noid, W. G., Krishna, V., Pfaendtner, J., and Voth, G. A. *Biophysical Journal* **95**(11), 5073 – 5083 (2008).

[194] Potestio, R., Fritsch, S., Español, P., Delgado-Buscalioni, R., Kremer, K., Everaers, R., and Donadio, D. *Physical Review Letters* **110**, 108301 (2013).

[195] Bucher, D., Hsu, Y.-H., Mouchlis, V. D., Dennis, E. A., and McCammon, J. A. *PLoS Computational Biology* **9**(7), e1003156 (2013).

[196] Espanol, P. In *Lecture notes of SoftSimu2002 - Novel Methods in Soft Matter Simulations,* Karttunen, M., Vattulainen, I., and Lukkarine, A., editors, Lecture Notes in Physics. Springer-Verlag (2003).

[197] Izvekov, S. *The Journal of Chemical Physics* **138**(13), 134106 (2013).

[198] Sinitskiy, A. V., Saunders, M. G., and Voth, G. A. *The Journal of Physical Chemistry B* **116**(29), 8363–8374 (2012).

[199] Guttenberg, N., Dama, J. F., Saunders, M. G., Voth, G. A., Weare, J., and Dinner, A. R. *The Journal of Chemical Physics* **138**(9), 094111 (2013).

[200] Shell, M. S. *The Journal of Chemical Physics* **129**(14), 144108 (2008).

[201] Henderson, R. *Physics Letters A* **49**(3), 197 – 198 (1974).

[202] Potestio, R. *Journal of Unsolved Questions* **3**(1), 13–15 (2013).

[203] Reith, D., Pütz, M., and Müller-Plathe, F. *Journal of Computational Chemistry* **24**(13), 1624–1636 (2003).

[204] Ercolessi, F. and Adams, J. B. *EPL (Europhysics Letters)* **26**(8), 583 (1994).

[205] Izvekov, S., Parrinello, M., Burnham, C. J., and Voth, G. A. *The Journal of Chemical Physics* **120**(23), 10896–10913 (2004).

[206] Izvekov, S. and Voth, G. A. *The Journal of Physical Chemistry B* **109**(7), 2469–2473 (2005).

[207] Lu, L., Dama, J. F., and Voth, G. A. *The Journal of Chemical Physics* **139**(12), 121906 (2013).

[208] Noid, W. G., Chu, J.-W., Ayton, G. S., Krishna, V., Izvekov, S., Voth, G. A., Das, A., and Andersen, H. C. *The Journal of Chemical Physics* **128**(24), 244114 (2008).

[209] Dama, J. F., Sinitskiy, A. V., McCullagh, M., Weare, J., Roux, B., Dinner, A. R., and Voth, G. A. *Journal of Chemical Theory and Computation* **9**(5), 2466–2480 (2013).

[210] Kusumaatmaja, H., Whittleston, C. S., and Wales, D. J. *Journal of Chemical Theory and Computation* **8**(12), 5159–5165 (2012).

[211] Deng, M. and Karniadakis, G. *Multiscale Modeling and Simulation* **12**(1), 109–118 (2014).

[212] Lyman, E., Pfaendtner, J., and Voth, G. A. *Biophysical Journal* **95**(9), 4183 – 4192 (2008).

[213] Sinitskiy, A. V. and Voth, G. A. *Chemical Physics* **422**(0), 165 – 174 (2013).

[214] Gur, M., Zomot, E., and Bahar, I. *The Journal of Chemical Physics* **139**(12), 121912 (2013).

[215] Amoros, D., Ortega, A., and Garcia de la Torre, J. *Journal of Chemical Theory and Computation* **9**(3), 1678–1685 (2013).

[216] Kim, Y. C. and Hummer, G. *Journal of Molecular Biology* **375**(5), 1416–1433 (2008).

[217] Rózycki, B., Kim, Y. C., and Hummer, G. *Structure* **19**(1), 109–116 (2011).

[218] Levine, A. J. *arXiv:cond-mat/0401624v1* (2004).

[219] Chakrabarti, B. and Levine, A. J. *Physical Review E* **74**, 031903 (2006).

[220] Chakrabarti, B. and Levine, A. J. *Physical Review E* **71**, 031905 (2005).

[221] Lee, P.-H., Helms, V., and Geyer, T. *The Journal of Chemical Physics* **137**(14), 145105 (2012).

[222] Zimmer, M. J. and Geyer, T. *The Journal of Chemical Physics* **136**(12), 125102 (2012).

[223] Chen, C., Wang, E., Liu, P., and Xiao, Y. *Physical Review E* **87**, 022701 (2013).

[224] Harding, J. H., Duffy, D. M., Sushko, M. L., Rodger, P. M., Quigley, D., and Elliott, J. A. *Chemical Reviews* **108**(11), 4823–4854 (2008).

[225] Heinz, H. *Journal of Computational Chemistry* **31**(7), 1564–1568 (2010).

[226] Feng, J., Pandey, R. B., Berry, R. J., Farmer, B. L., Naik, R. R., and Heinz, H. *Soft Matter* **7**, – (2011).

[227] Utesch, T., Daminelli, G., and Mroginski, M. A. *Langmuir* **27**(21), 13144–13153 (2011).

[228] Cormack, A. N., Lewis, R. J., and Goldstein, A. H. *The Journal of Physical Chemistry B* **108**(52), 20408–20418 (2004).

[229] Song, D. and Forciniti, D. *The Journal of Chemical Physics* **115**(17), 8089–8100 (2001).

[230] Utesch, T., Sezer, M., Weidinger, I. M., and Mroginski, M. A. *Langmuir* **28(13)**, 5761 (2012).

[231] Ganazzoli, F. and Raffaini, G. *Physical Chemistry Chemical Physics* **7**, 3651–3663 (2005).

[232] Agashe, M., Raut, V., Stuart, S. J., and Latour, R. A. *Langmuir* **21**(3), 1103–1117 (2005).

[233] Raffaini, G. and Ganazzoli, F. *The Journal of Physical Chemistry B* **108**(36), 13850–13854 (2004).

[234] Raffaini, G. and Ganazzoli, F. *Langmuir* **19**(8), 3403–3412 (2003).

[235] Raffaini, G. and Ganazzoli, F. *Langmuir* **20**(8), 3371–3378 (2004).

[236] Raffaini, G. and Ganazzoli, F. *Journal of Biomedical Materials Research Part A* **76A**(3), 638–645 (2006).

[237] Hähl, H., Evers, F., Grandthyll, S., Paulus, M., Sternemann, C., Loskill, P., Lessel, M., Hüsecken, A. K., Brenner, T., Tolan, M., and Jacobs, K. *Langmuir* **28**(20), 7747–7756 (2012).

[238] Kang, S.-g., Huynh, T., Xia, Z., Zhang, Y., Fang, H., Wei, G., and Zhou, R. *Journal of the American Chemical Society* **135**(8), 3150–3157 (2013).

[239] Bachmann, M., Goede, K., Beck-Sickinger, A. G., Grundmann, M., Irbäck, A., and Janke, W. *Angewandte Chemie International Edition* **49**(49), 9530–9533 (2010).

[240] Patwardhan, S. V., Emami, F. S., Berry, R. J., Jones, S. E., Naik, R. R., Deschaume, O., Heinz, H., and Perry, C. C. *Journal of the American Chemical Society* **134**(14), 6244–6256 (2012).

[241] Emami, F. S., Puddu, V., Berry, R. J., Varshney, V., Patwardhan, S. V., Perry, C. C., and Heinz, H. *Chemistry of Materials* **26**(19), 5725–5734 (2014).

[242] Herrera, F. E., Bouchet, A., Lairion, F., Disalvo, E. A., and Pantano, S. *The Journal of Physical Chemistry B* **116**(15), 4476–4483 (2012).

[243] Siwko, M. and Corni, S. *Physical Chemistry Chemical Physics* **15**, 5945–5956 (2013).

[244] Fackovec, B. and Vondrasek, J. S. *Journal of Physical Chemistry B* **116**(42), 12651–126660 (2012).

[245] Raffaini, G. and Ganazzoli, F. *Macromolecular Bioscience* **7**(5), 552–566 (2007).

[246] Bajd, F. and Sersa, I. *Biophysical Journal* **104**(5), 1181 – 1190 (2013).

[247] Weisel, J. and Nagaswami, C. *Biophysical Journal* **63**(1), 111 – 128 (1992).

[248] Magatti, D., Molteni, M., Cardinali, B., Rocco, M., and Ferri, F. *Biophysical Journal* **104**(5), 1151 – 1159 (2013).

[249] Lim, B. B., Lee, E. H., Sotomayor, M., and Schulten, K. *Structure* **16**(3), 449–459 (2008).

[250] Zhmurov, A., Brown, A. E., Litvinov, R. I., Dima, R. I., Weisel, J. W., and Barsegov, V. *Structure* **19**(11), 1615 (2011).

[251] Hyeon, C., Dima, R. I., and Thirumalai, D. *Structure* **14**(11), 1633–1645 (2006).

[252] Kononova, O., Litvinov, R. I., Zhmurov, A., Alekseenko, A., Cheng, C. H., Agarwal, S., Marx, K. A., Weisel, J. W., and Barsegov, V. *The Journal of Biological Chemistry* **288**(31), 22681–22692 (2013).

[253] Averett, R. D., Menn, B., Lee, E. H., Helms, C. C., Barker, T., and Guthold, M. *Biophysical Journal* **103**(7), 1537–1544 (2012).

[254] Adamczyk, Z., Barbasz, J., and Cieśla, M. *Langmuir* **26**(14), 11934–11945 (2010).

[255] Adamczyk, Z., Barbasz, J., and Cieśla, M. *Langmuir* **27**(11), 6868–6878 (2011).

[256] Vilaseca, P., Dawson, K. A., and Franzese, G. *Soft Matter* **9**, 6978–6985 (2013).

[257] Dabkowska, M. and Adamczyk, Z. *Langmuir* **28**(44), 15663–15673 (2012).

[258] Ciesla, M. and Barbasz, J. In *ISSIS 2012 conference*, (2012). arXiv:1208.0167.

[259] www.paramchem.org.

[260] Humphrey, W., Dalke, A., and Schulten, K. *Journal of Molecular Graphics* **14**, 33 (1996).

[261] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. *Journal of Computational Chemistry* **26**, 1781 (2005).

[262] Kilpatrick, J. I., Loh, S.-H., and Jarvis, S. P. *Journal of the American Chemical Society* **135**(7), 2628–2634 (2013).

[263] Gordillo, M. C. and Marti, J. *The Journal of Chemical Physics* **117**(7), 3425–3430 (2002).

[264] Werder, T., Walther, J. H., Jaffe, R. L., Halicioglu, T., and Koumoutsakos, P. *The Journal of Physical Chemistry B* **107**(6), 1345–1352 (2003).

[265] Ma, Q., Izaguirre, J. A., and Skeel, R. D. *SIAM Journal on Scientific Computing* **24**(6), 1951–1973 (2003).

[266] Martyna, G. J., Tobias, D. J., and Klein, M. L. *The Journal of Chemical Physics* **101**(5), 4177–4189 (1994).

[267] Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. *The Journal of Chemical Physics* **103**(11), 4613–4621 (1995).

[268] Kitao, A., Hirata, F., and Go, N. *Chemical Physics* **158**(2–3), 447 − 472 (1991).

[269] Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. *Proteins: Structure, Function, and Bioinformatics* **17**(4), 412–425 (1993).

[270] Hess, B. *Physical Review E* **62**, 8438–8448 (2000).

[271] Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caflisch, A. *Bioinformatics* **23**(19), 2625–2627 (2007).

[272] Bakan, A., Meireles, L. M., and Bahar, I. *Bioinformatics* **27**(11), 1575–1577 (2011).

[273] Lange, O. F. and Grubmüller, H. *Proteins* **62**(4), 1053–1061 (2006).

[274] Ribeiro, A. A. S. T. and Ortiz, V. *Journal of Chemical Theory and Computation* **10**(4), 1762–1769 (2014).

[275] Lange, O. F. and Grubmüller, H. *Proteins: Structure, Function, and Bioinformatics* **70**(4), 1294–1312 (2008).

[276] Bradley, M. J., Chivers, P. T., and Baker, N. A. *Journal of Molecular Biology* **378**(5), 1155–1173 (2008).

[277] Papaleo, E., Lindorff-Larsen, K., and De Gioia, L. *Physical Chemistry Chemical Physics* **14**, 12515–12525 (2012).

[278] Chiappori, F., Merelli, I., Colombo, G., Milanesi, L., and Morra, G. *PLoS Computational Biology* **8**(12), e1002844 (2012).

[279] Feher, V. A., Durrant, J. D., Wart, A. T. V., and Amaro, R. E. *Current Opinion in Structural Biology* **25C**, 98–103 (2014).

[280] Kabsch, W. and Sander, C. *Biopolymers* **22**(12), 2577–2637 (1983).

[281] Hekkelman, M. DSSP 2.0 (http://www.cmbi.ru.nl/dssp.html).

[282] Poornam, G. P., Matsumoto, A., Ishida, H., and Hayward, S. *Proteins: Structure, Function, and Bioinformatics* **76**(1), 201–212 (2009).

[283] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **98**(18), 10037–10041 (2001).

[284] Holst, M. and Saied, F. *Journal of Computational Chemistry* **14**(1), 105–113 (1993).

[285] UniProt Consortium. *Nucleic Acids Research* **42**(Database issue), D191–D198 (2014).

[286] www.uniprot.org.

[287] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. *Journal of Computational Chemistry* **25**(13), 1605–1612 (2004).

[288] Meng, E., Pettersen, E., Couch, G., Huang, C., and Ferrin, T. *BMC Bioinformatics* **7**(1), 339 (2006).

[289] Lacroix, E., Viguera, A. R., and Serrano, L. *Journal of Molecular Biology* **284**(1), 173 – 191 (1998).

[290] Munoz, V. and Serrano, L. *Nature structural and molecular biology* **1**, 399 (1994).

[291] http://agadir.crg.es/.

[292] Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. *Structure* **11**(11), 1453 – 1459 (2003).

[293] Parrinello, M. and Rahman, A. *Journal of Applied Physics* **52**(12), 7182–7190 (1981).

[294] Hess, B. *Physical Review E* **65**, 031910 (2002).

[295] Huse, D. A. and Henley, C. L. *Physical Review Letters* **54**, 2708–2711 (1985).

[296] Hanss, M. and Biot, F. *Annals of the New York Academy of Sciences* **936**, 89–90 (2001). http://site.geht.org/pages/?page=40&idl=21.

[297] Asselta, R., Duga, S., Spena, S., Peyvandi, F., Castaman, G., Malcovati, M., Mannucci, P. M., and Tenchini, M. L. *Blood* **103**(8), 3051–3054 (2004).

[298] Maghzal, G. J., Brennan, S. O., Fellowes, A. P., Spearing, R., and George, P. M. *Biochimica et Biophysica Acta* **1645**(2), 146–151 (2003).

[299] Sugo, T., Nakamikawa, C., Takano, H., Mimuro, J., Yamaguchi, S., Mosesson, M. W., Meh, D. A., DiOrio, J. P., Takahashi, N., Takahashi, H., Nagai, K., and Matsuda, M. *Blood* **94**(11), 3806–3813 (1999).

[300] Margaglione, M., Santacroce, R., Colaizzo, D., Seripa, D., Vecchione, G., Lupone, M. R., De Lucia, D., Fortina, P., Grandone, E., Perricone, C., and Di Minno, G. *Blood* **96**(7), 2501–2505 (2000).

[301] Terasawa, F., Kamijyo, Y., Fujihara, N., Yamauchi, K., Kumagai, T., Honda, T., Shigematsu, S., and Okumura, N. *Clinica Chimica Acta* **411**(17-18), 1325–1329 (2010).

[302] Brennan, S. O., Fellowes, A. P., Faed, J. M., and George, P. M. *Blood* **95**(5), 1709–1713 (2000).

[303] Zdziarska, J., Undas, A., Basa, J., Iwaniec, T abd Skotnicki, A., de Moerloose, P., and Neerman-Arbez, M. *Blood coagulation & fibrinolysis* **20**(5), 374–6 (2009).

[304] Ivaskevicius, V., Jusciute, E., Steffens, M., Geisen, C., Hanfland, P., Wienker, T. F., Seifried, E., and Oldenburg, J. *Blood Coagulation & Fibrinolysis* **16**(3), 205–208 (2005).

[305] Brennan, S. O., Homer, V. M., Davis, R. L., Meyer, M., and George, P. M. *Thrombosis and Haemostasis* **96**(4), 535–537 (2006).

[306] Maekawa, H., Yamazumi, K., Muramatsu, S., Kaneko, M., Hirata, H., Takahashi, N., Arocha-Pinango, C. L., Rodriguez, S., Nagy, H., and Perez-Requejo, J. L. *The Journal of Clinical Investigation* **90**(1), 67–76 (1992).

[307] Plate, M., Asselta, R., Peyvandi, F., Tenchini, M. L., and Duga, S. *Biochimica et Biophysica Acta* **1772**(7), 781 – 787 (2007).

[308] Platè, M., Asselta, R., Spena, S., Spreafico, M., Fagoonee, S., Peyvandi, F., Tenchini, M. L., and Duga, S. *Blood Cells, Molecules and Diseases* **41**(3), 292–297 (2008).

[309] Marchi, R. C., Meyer, M. H., de Bosch, N. B., Arocha-Pinango, C. L., and Weisel, J. W. *Blood Coagulation & Fibrinolysis* **15**(7), 559–567 (2004).

[310] Morris, T. A., Marsh, J. J., Chiles, P. G., Magaña, M. M., Liang, N.-C., Soler, X., Desantis, D. J., Ngo, D., and Woods, Jr, V. L. *Blood* **114**(9), 1929–1936 (2009).

[311] Okumura, N., Terasawa, F., Hirota-Kawadobora, M., Yamauchi, K., Nakanishi, K., Shiga, S., Ichiyama, S., Saito, M., Kawai, M., and Nakahata, T. *Clinica Chimica Acta* **365**(1-2), 160–167 (2006).

[312] Hanss, M., Ffrench, P., Viniciguerra, C., Bertrand, M.-A., and De Mazancourt, P. *Journal of Thrombosis and Haemostasis* **3**(10), 2347–2349 (2005).

[313] Lounes, K. C., Lefkowitz, J. B., Coates, A. I., Hantgan, R. R., Henschen-Edman, A., and Lord, S. T. *Annals of the New York Academy of Sciences* **936**, 129–132 (2001).

[314] Lounes, K. C., Lefkowitz, J. B., Henschen-Edman, A. H., Coates, A. I., Hantgan, R. R., and Lord, S. T. *Blood* **98**(3), 661–666 (2001).

[315] Brown, J. H., Volkmann, N., Jun, G., Henschen-Edman, A. H., and Cohen, C. *Proceedings of the National Academy of Sciences of the United States of America* **97**(1), 85–90 (2000).

[316] Yang, Z., Kollman, J. M., Pandi, L., and Doolittle, R. F. *Biochemistry* **40**(42), 12515–12523 (2001).

[317] Monaldini, L., Asselta, R., Duga, S., Peyvandi, F., Ghosh, K., Malcovati, M., and Tenchini, M. L. *Haematologica* **91**(5), 628–633 (2006).

[318] Fan, J., Saunders, M., and Voth, G. *Biophysical Journal* **103**(6), 1334 – 1342 (2012).

[319] Everse, S. J., Spraggon, G., and Doolittle, R. F. *Thrombosis and Haemostasis* **80**(1), 1–9 (1998).

[320] Miyata, T., Furukawa, K., Iwanaga, S., Takamatsu, J., and Saito, H. *The Journal of Biochemistry* **105**(1), 10–14 Jan (1989).

[321] Terukina, S., Yamazumi, K., Okamoto, K., Yamashita, H., Ito, Y., and Matsuda, M. *Blood* **74**(8), 2681–2687 (1989).

[322] Reber, P., Furlan, M., Rupp, C., Kehl, M., Henschen, A., Mannucci, P. M., and Beck, E. A. *Blood* **67**(6), 1751–1756 (1986).

[323] Lounes, K. C., Soria, C., Mirshahi, S. S., Desvignes, P., Mirshahi, M., Bertrand, O., Bonnet, P., Koopman, J., and Soria, J. *Blood* **96**(10), 3473–3479 (2000).

[324] Okumura, N., Furihata, K., Terasawa, F., Nakagoshi, R., Ueno, I., and Katsuyama, T. *Thrombosis and Haemostasis* **75**(6), 887–891 (1996).

[325] Bentolila, S., Samama, M. M., Conard, J., Horellou, M. H., and Ffrench, P. *Annales de Médecine Interne* **146**(8), 575–580 (1995).

[326] Brennan, S. O., Maghzal, G., Shneider, B. L., Gordon, R., Magid, M. S., and George, P. M. *Hepatology* **36**(3), 652–658 (2002).

[327] Rubbia-Brandt, L., Neerman-Arbez, M., Rougemont, A.-L., Malé, P.-J., and Spahr, L. *The American Journal of Surgical Pathology* **30**(7), 906–911 (2006).

[328] Sciveres, M., Francalanci, P., Caprai, S., Talini, I., Callea, F., and Maggiore, G. *Journal of Pediatric Gastroenterology & Nutrition* **42**(5), E75–E76 (2006).

[329] Francalanci, P., Santorelli, F. M., Talini, I., Boldrini, R., Devito, R., Camassei, F. D., Maggiore, G., and Callea, F. *The Journal of Pediatrics* **148**(3), 396 – 398 (2006).

[330] Sogo, T., Nagasaka, H., Komatsu, H., Inui, A., Miida, T., Callea, F., Francalanci, P., Hirano, K.-I., Kitamura, H., Yorifuji, T., and Fujisawa, T. *Journal of Pediatric Gastroenterology & Nutrition* **49**(1), 133–136 (2009).

[331] Al-Hussaini, A., Altalhi, A., El Hag, I., AlHussaini, H., Francalanci, P., Giovannoni, I., and Callea, F. *Saudi Journal of Gastroenterology* **20**(4), 255–261 (2014).

[332] Yoshida, N., Hirata, H., Morigami, Y., Imaoka, S., Matsuda, M., Yamazumi, K., and Asakura, S. *The Journal of Biological Chemistry* **267**(4), 2753–2759 (1992).

[333] Litvinov, R. I., Gorkun, O. V., Galanakis, D. K., Yakovlev, S., Medved, L., Shuman, H., and Weisel, J. W. *Blood* **109**(1), 130–138 (2007).

[334] Davis, R. L., Mosesson, M. W., Kerlin, B. A., Canner, J. A., Ruymann, F. B., and Brennan, S. O. *Haematologica* **92**(8), 1151–1152 (2007).

[335] Hanss, M., Chevreaud, C., French, P., Negrier, C., and de Mazancourt, P. *Thrombosis and Haemostasis* **98**(3), 689–691 (2007).

[336] Hamano, A., Mimuro, J., Aoshima, M., Itoh, T., Kitamura, N., Nishinarita, S., Takano, K., Ishiwata, A., Kashiwakura, Y., Niwa, K., Ono, T., Madoiwa, S., Sugo, T., Matsuda, M., and Sakata, Y. *Blood* **103**(8), 3045–3050 (2004).

[337] Martinez, J., Holburn, R. R., Shapiro, S. S., and Erslev, A. J. *Journal of Clinical Investigation* **53**(2), 600–611 (1974).

[338] Keller, M. A., Martinez, J., Baradet, T. C., Nagaswami, C., Chernysh, I. N., Borowski, M. K., Surrey, S., and Weisel, J. W. *Blood* **105**(8), 3162–3168 (2005).

[339] Ripka, L. Master's thesis, Johannes Gutenberg-Universität Mainz, (2014).

[340] Zhang, Z., Pfaendtner, J., Grafmüller, A., and Voth, G. A. *Biophysical Journal* **97**(8), 2327 – 2337 (2009).

[341] Zhang, Z. and Voth, G. A. *Journal of Chemical Theory and Computation* **6**(9), 2990–3002 (2010).

[342] Van Wynsberghe, A. W. and Cui, Q. *Structure* **14**(11), 1647–1653 (2006).

[343] Grime, J. M. A. and Voth, G. A. *Biophysical Journal* **103**(8), 1774–1783 (2012).

[344] Arkhipov, A., Freddolino, P. L., and Schulten, K. *Structure* **14**(12), 1767–1777 Dec (2006).

[345] Ping, L., Huang, L., Cardinali, B., Profumo, A., Gorkun, O. V., and Lord, S. T. *Biochemistry* **50**(42), 9066–9075 (2011).

[346] Köhler, S., Schmid, F., and Settanni, G. *submitted* (2015).

[347] Laio, A. and Parrinello, M. *Proceedings of the National Academy of Sciences of the United States of America* **99**(20), 12562–12566 (2002).

[348] Bussi, G., Laio, A., and Parrinello, M. *Physical Review Letters* **96**(9), 090601 (2006).

[349] Litvinov, R. I. and Weisel, J. W. *Journal of Thrombosis and Haemostasis* **11**(10), 1933–1935 (2013).

[350] Köhler, S., Ruocco, G., and Schirmacher, W. *Physical Review B: Condensed Matter and Materials Physics* **88**, 064203 (2013).

[351] de Leeuw, S. W., Perram, J. W., and Smith, E. R. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **373**(1752), 27–56 (1980).

[352] W.Rocchia, E.Alexov, and B.Honig. *The Journal of Physical Chemistry B* **105**(28), 6507–6514 (2001).

[353] Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. *Journal of Computational Chemistry* **23**(1), 128–137 (2002).

Part VI

ATTACHMENTS

PREVIOUS PUBLICATIONS

Some results of the research in this thesis have been presented in the following publications:

1. S. Köhler and G. Settanni, "The internal dynamics of fibrinogen and its implications for coagulation and adsorption", *European Biophysical Journal* **42** (Suppl. 1), S92 (2013).

2. S. Köhler, F. Schmid, and G. Settanni, "The Flexibility of Fibrinogen and its initial Adsorption Stages at the Graphite and Mica Surface", In *Proceedings of the NIC Symposium 2014*, K. Binder, G. Münster, M. Kremer (Editors), NIC Series Vol. 47, 117–125 (2014)

3. S. Köhler, F. Schmid, and G. Settanni, "The flexibility of fibrinogen and its implications for binding and adsorption", PLoS Computational Biology, **in Press** (2015).